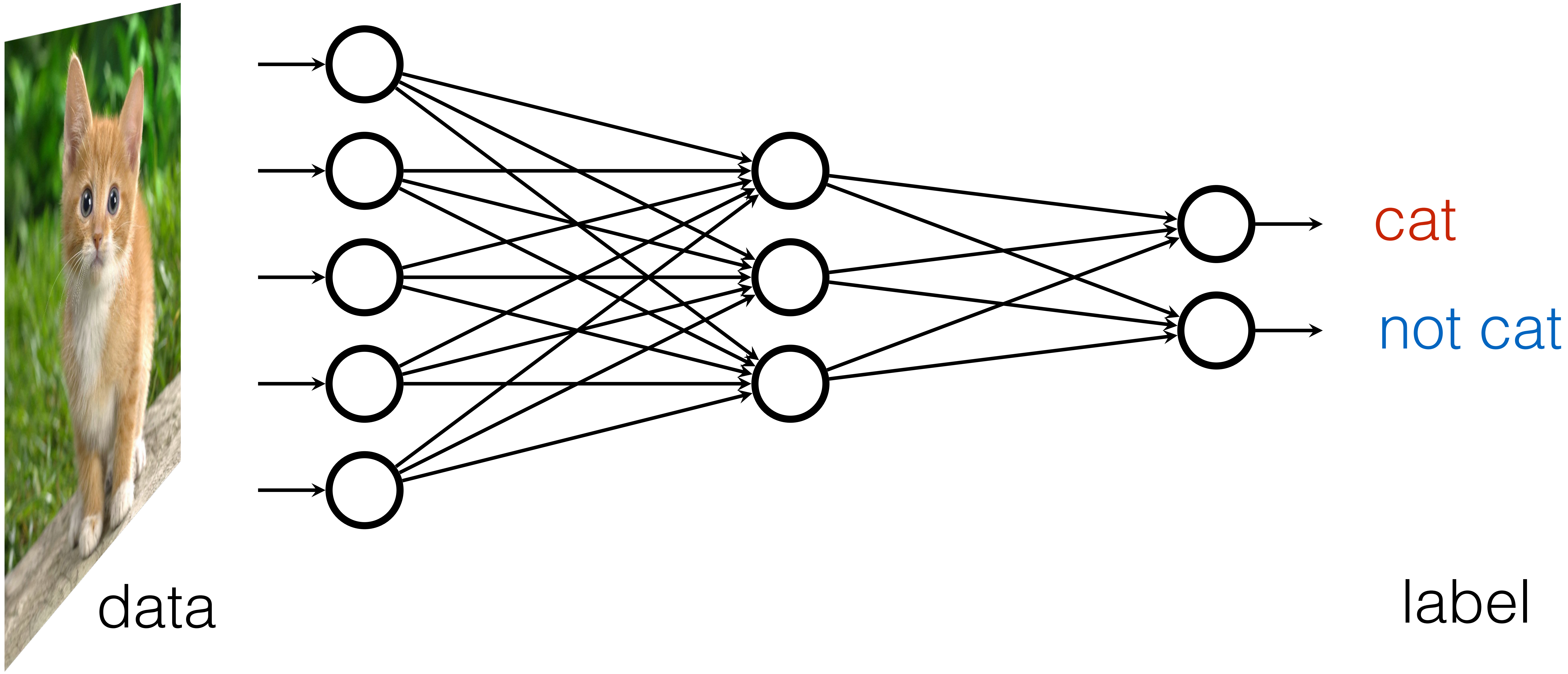
From Boltzmann Machines to Born Machines

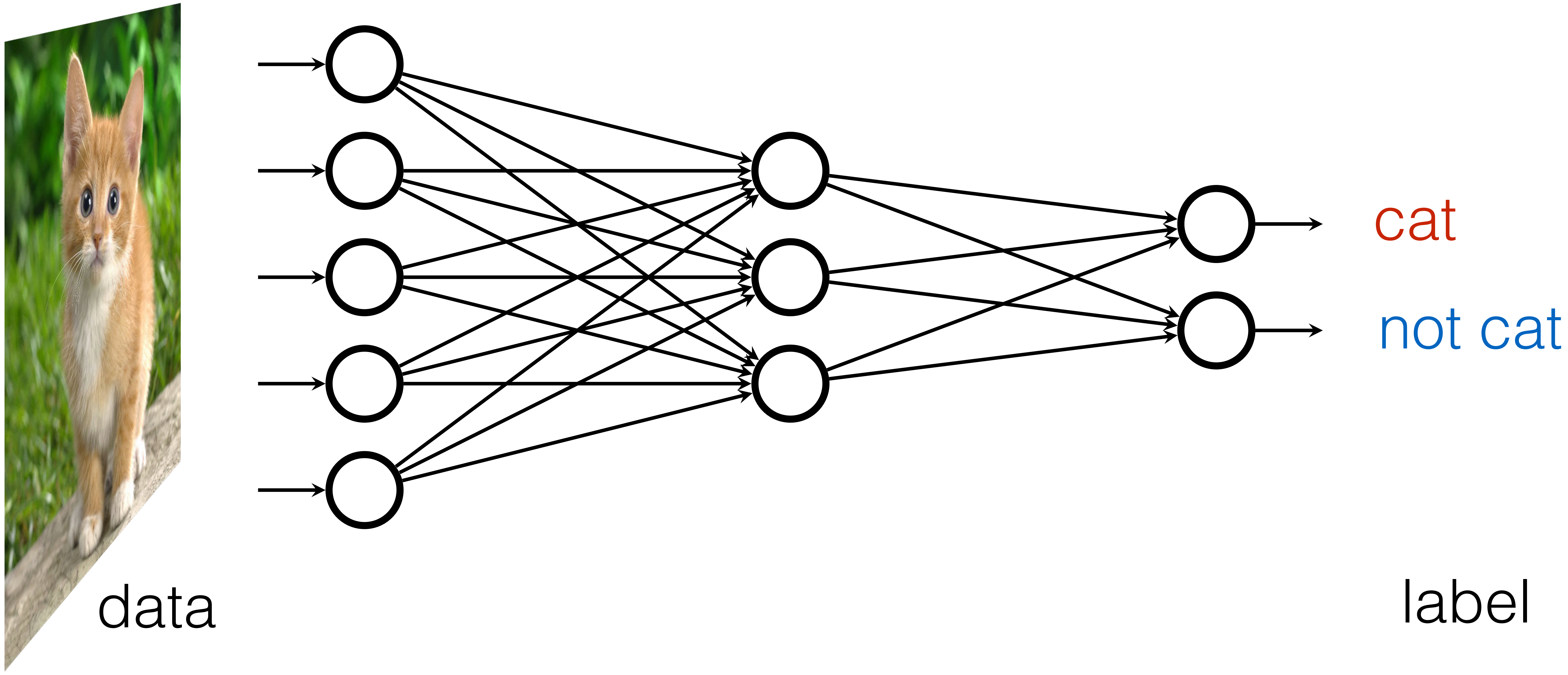
Lei Wang (王磊)
https://wangleiphy.github.io

Institute of Physics, Beijing
Chinese Academy of Sciences

Pattern recognition and beyond



Pattern recognition and beyond

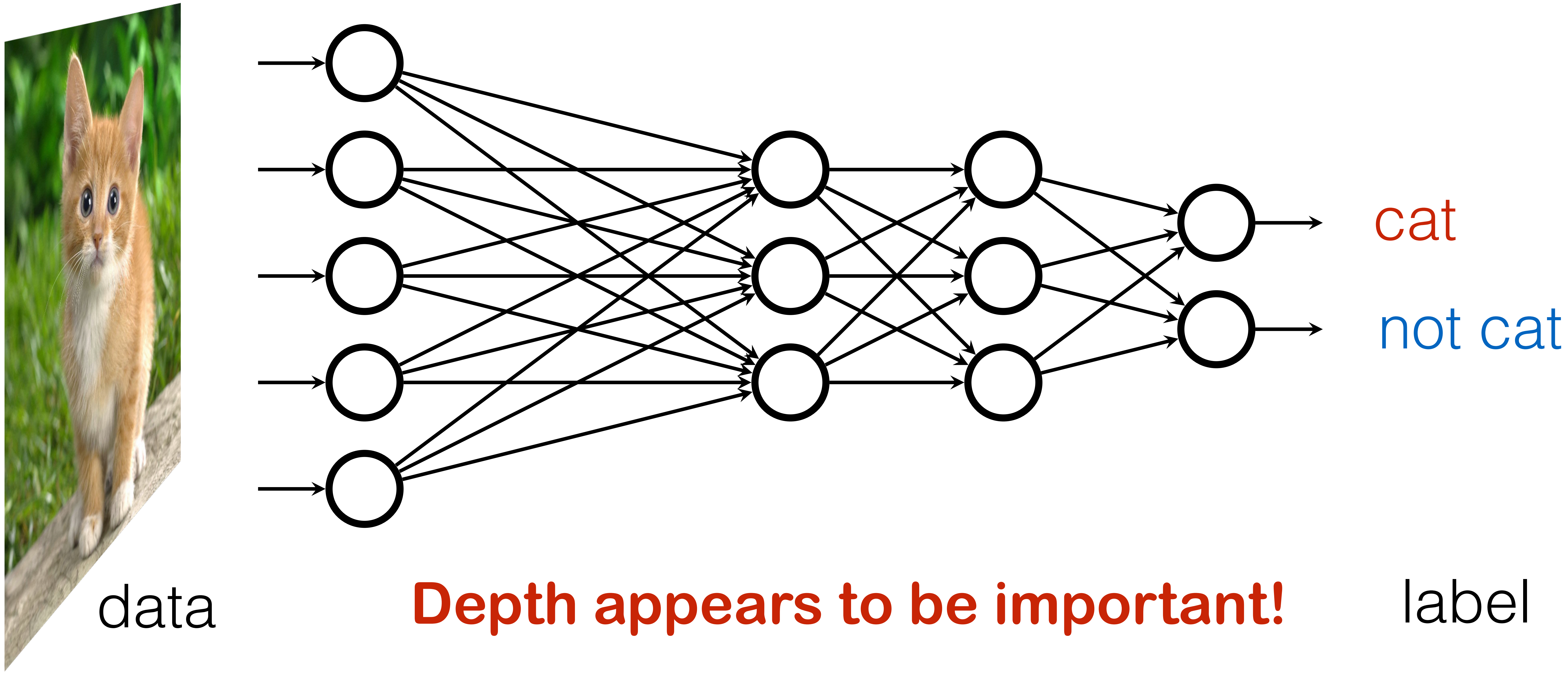


Universal Function Approximator

Cybenko 1989

Hornik, Stinchcombe, White 1989

Pattern recognition and beyond

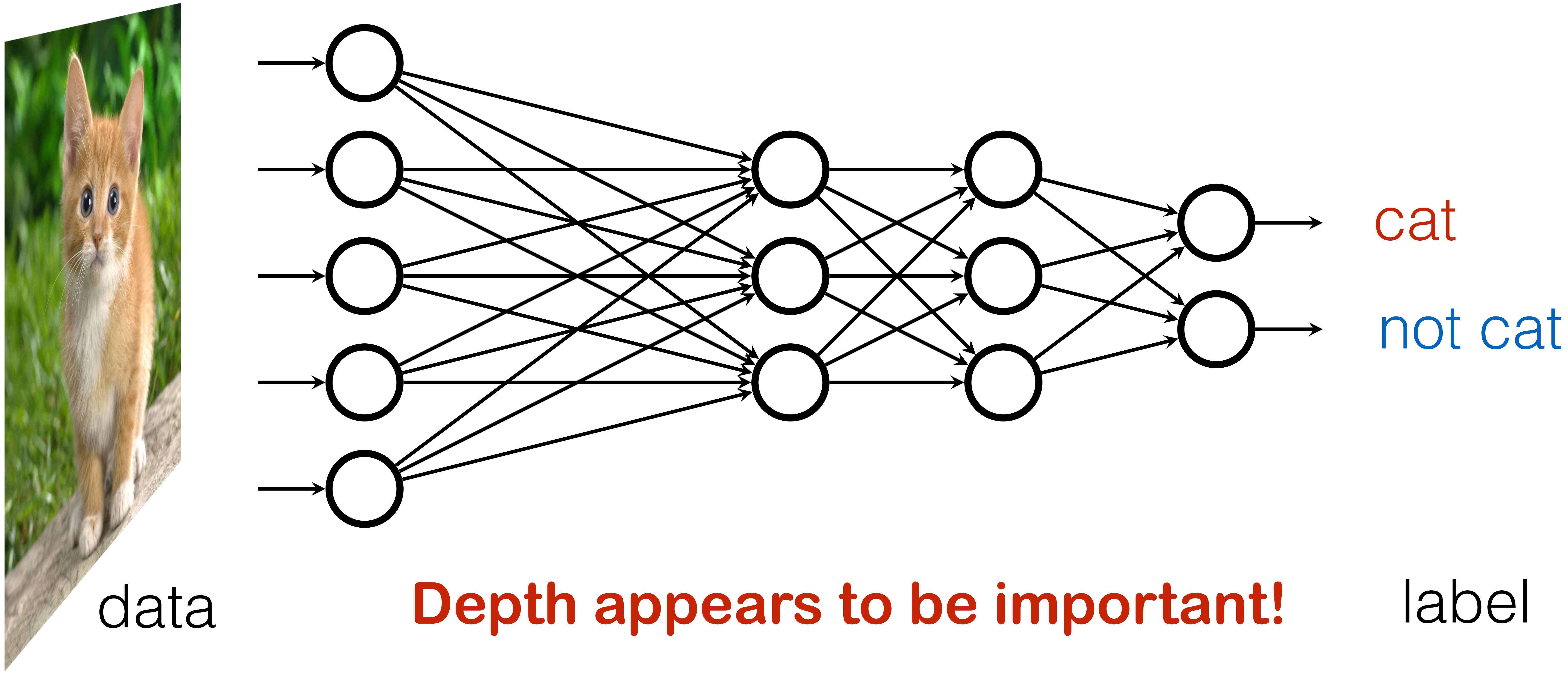


data

Depth appears to be important!

label

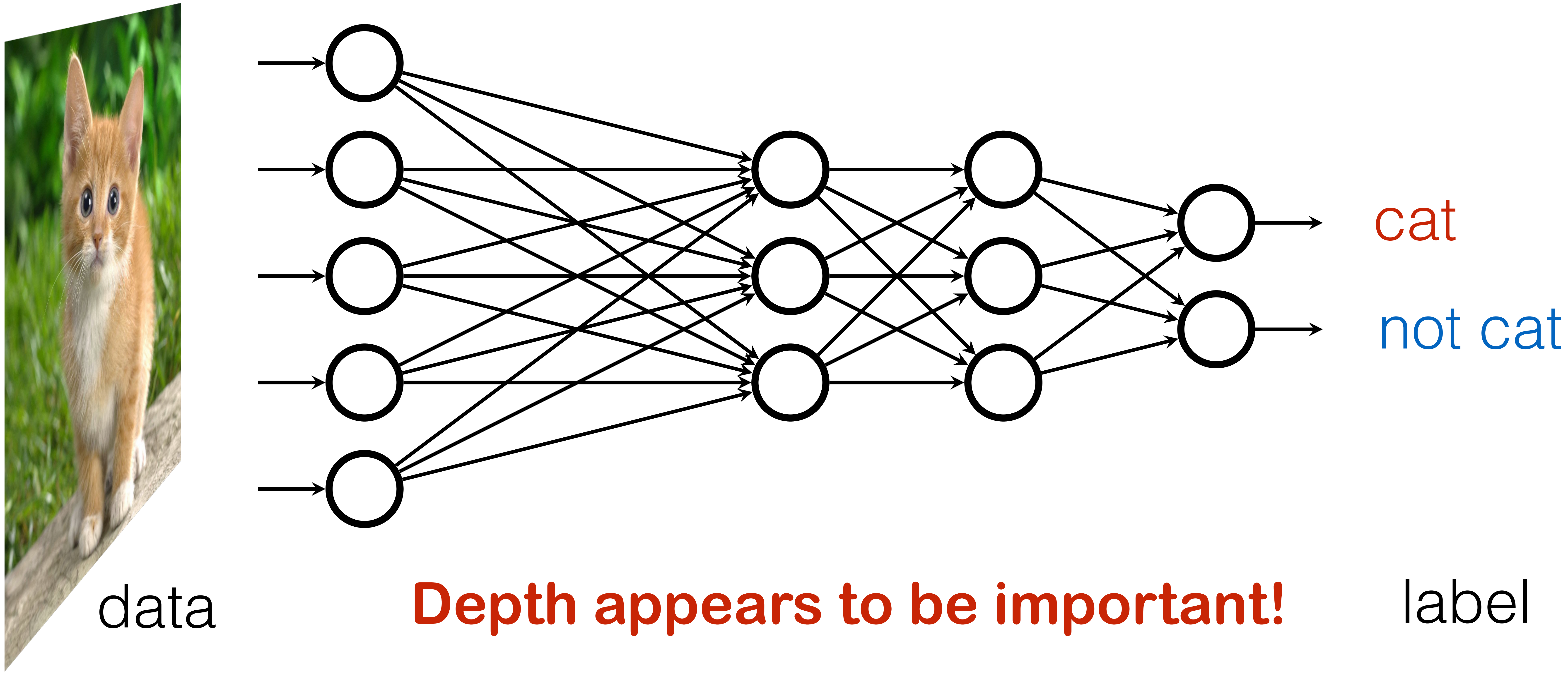
Pattern recognition and beyond



Depth appears to be important!

Q: Why does deep learning work?

Pattern recognition and beyond



Depth appears to be important!

Q: Why does deep learning work?

A: Law of physics: symmetry, locality, compositionality, renormalization group, and quantum entanglement.

Bény, 1301.3124 Mehta, Schwab,1410.3831 Lin, Tegmark, Rolnick,1608.08225 Levine et al, 1704.01552 …

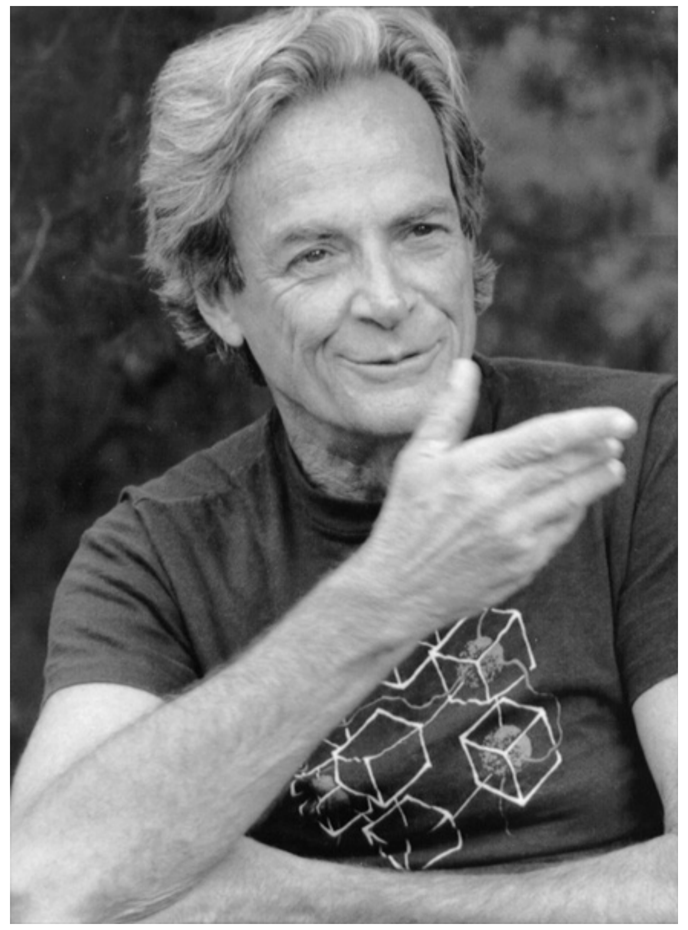
Discriminative vs Generative Learning

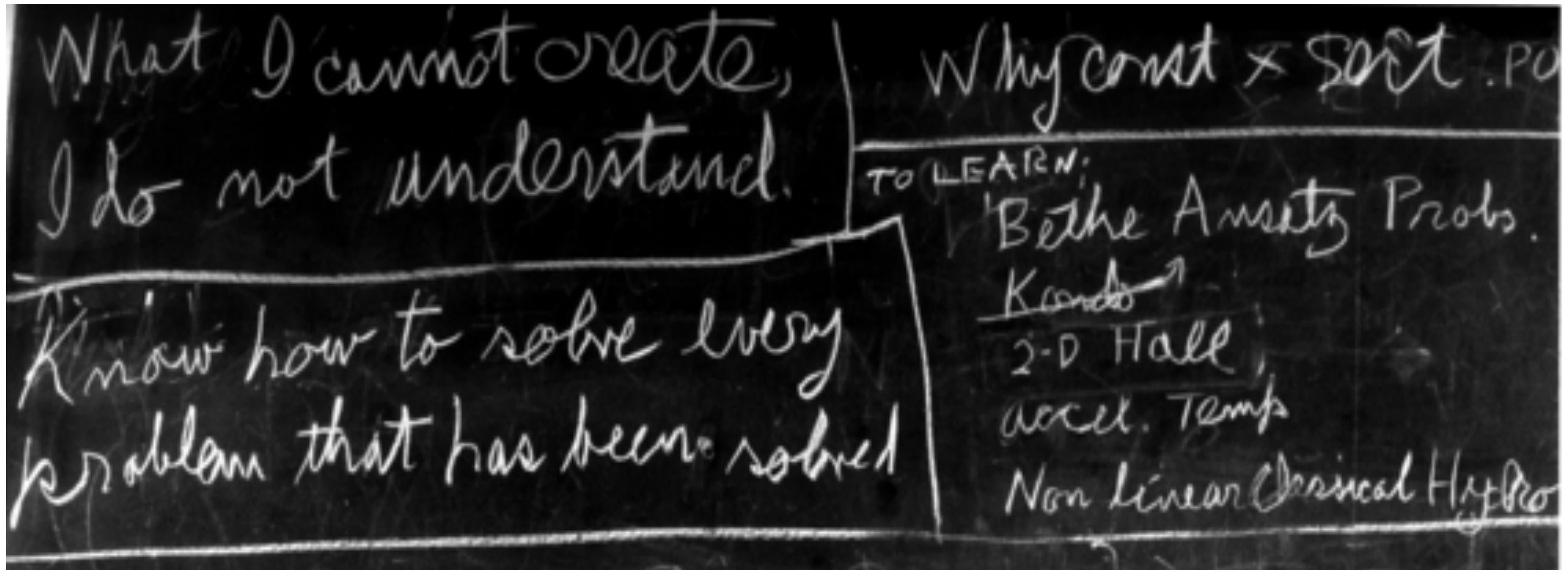
Discriminative vs Generative Learning

read

write

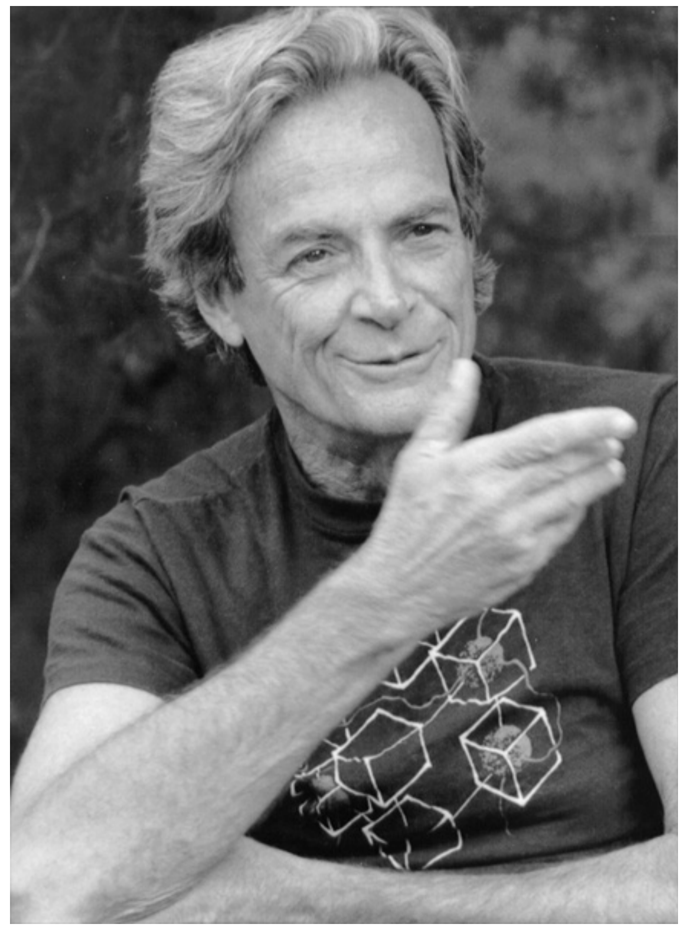
Discriminative vs Generative Learning

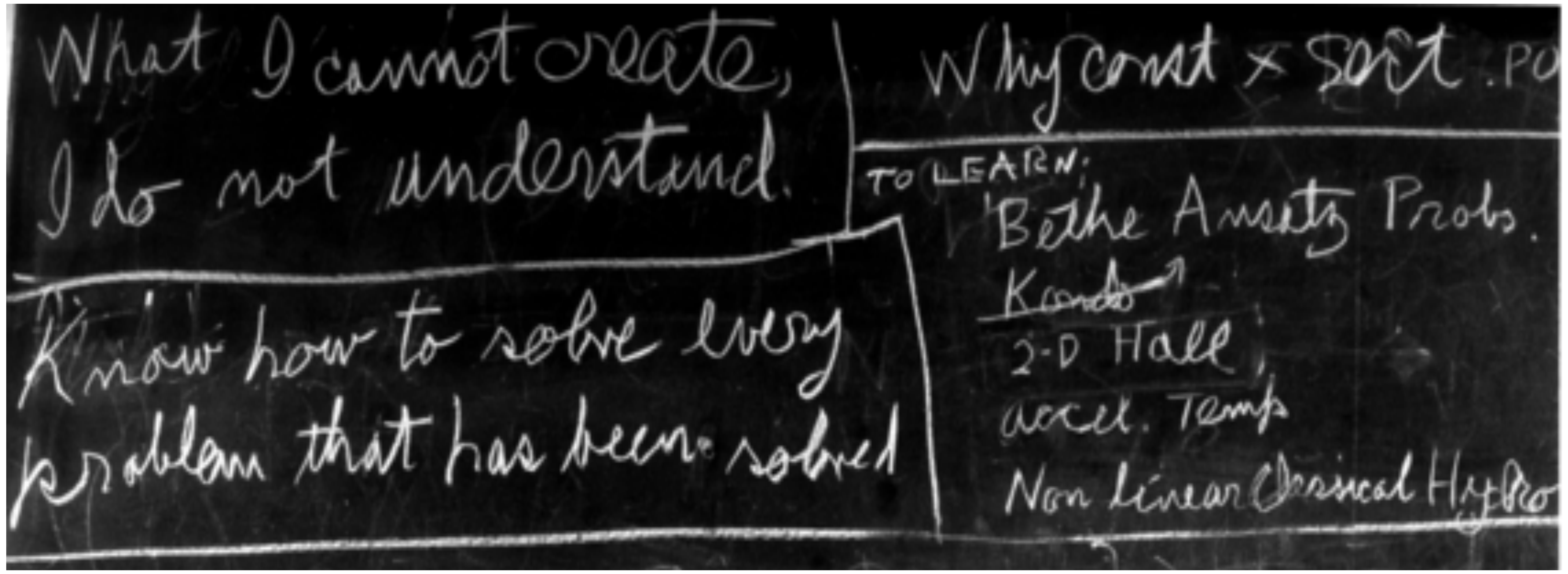




“What I can not create, I do not understand”

Discriminative vs Generative Learning





ELSEVIER

Progress in Brain Research

Volume 165, 2007, Pages 535–547

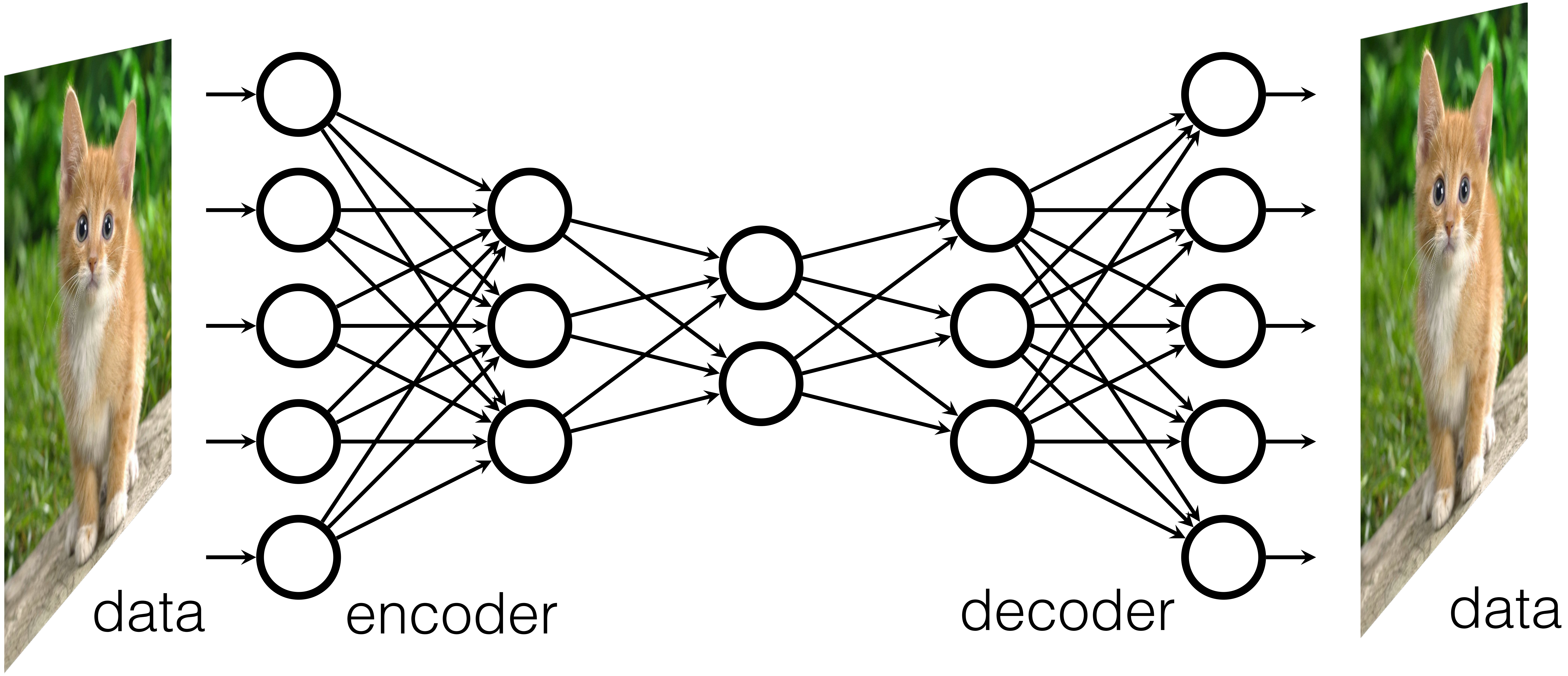
Computational Neuroscience: Theoretical Insights into Brain Function

To recognize shapes, first learn to generate images

Geoffrey E. Hinton

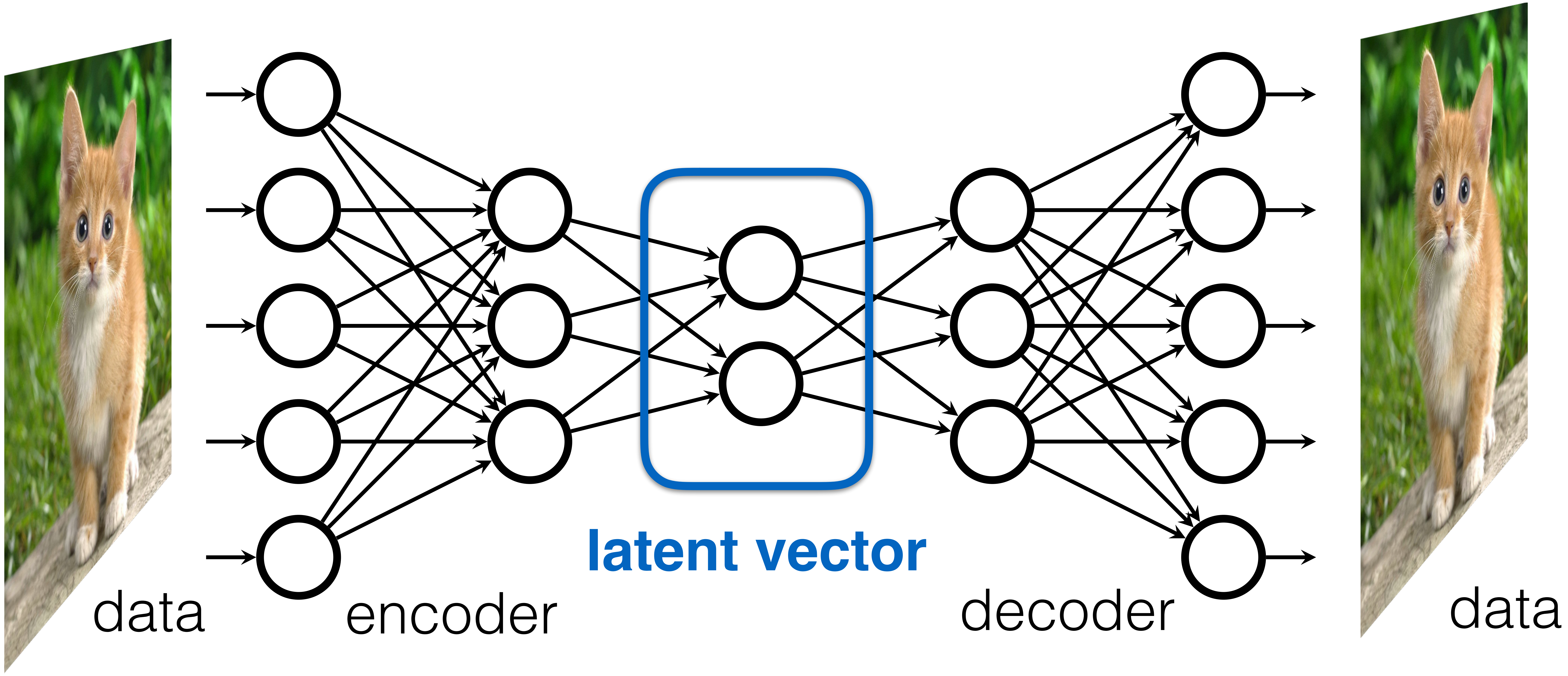
Department of Computer Science, University of Toronto, 10 Kings College Road, Toronto, M5S 3G4 Canada

Generative Modeling



“Auto-Encoding Variational Bayes", Kingma and Welling,1312.6114

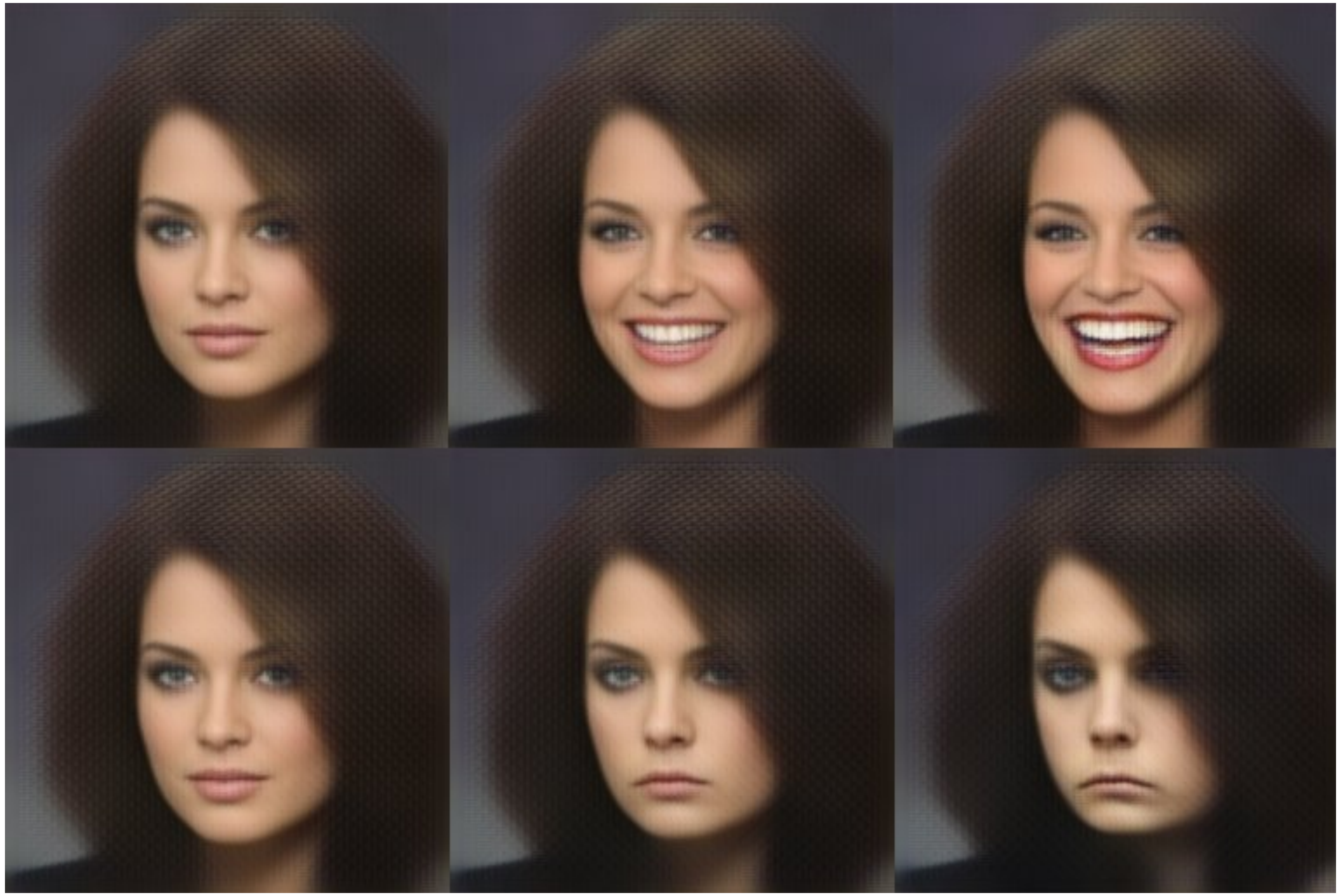
Generative Modeling



"Auto-Encoding Variational Bayes", Kingma and Welling,1312.6114

Interpolation in the latent space

arithmetics of the “smile vector”

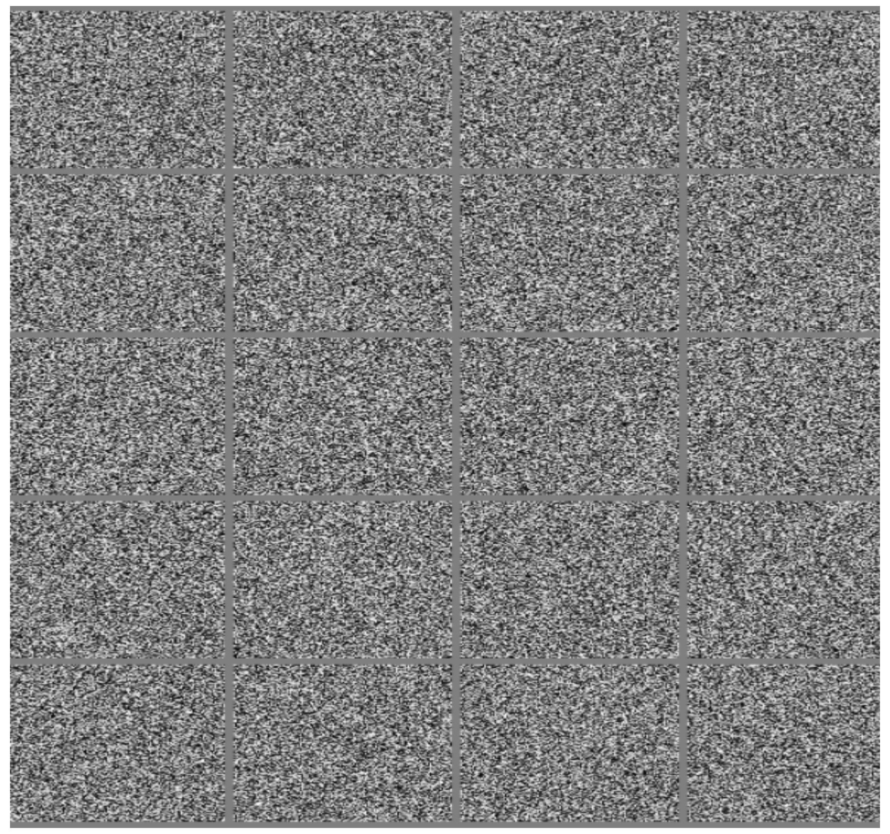


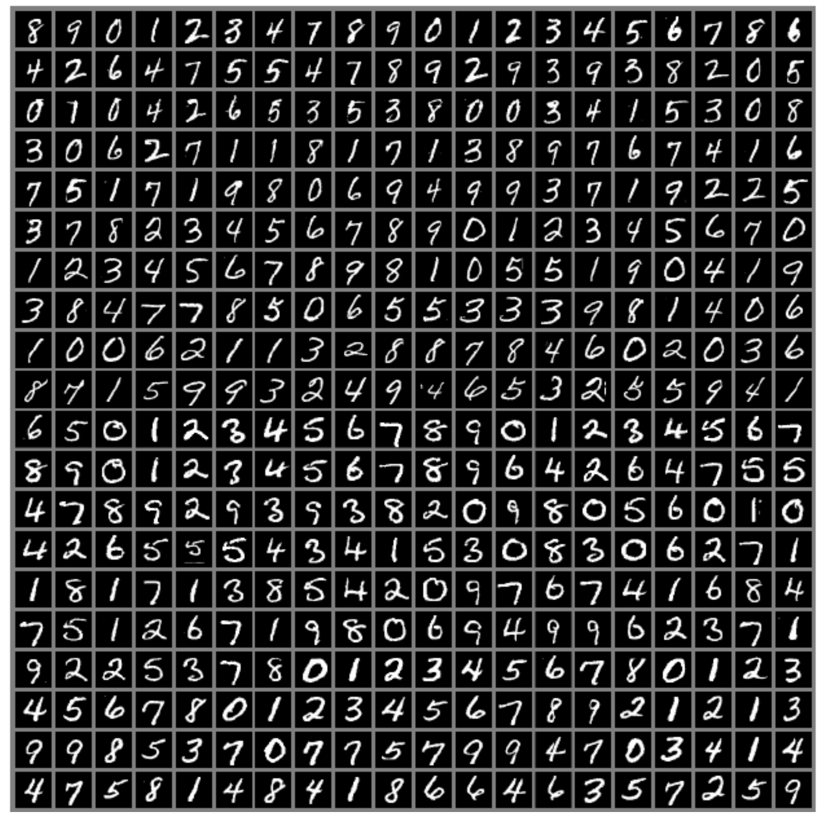
White, 1609.04468

Probabilistic Generative Modeling

$$p(\boldsymbol{x})$$

How to express, learn, and sample from a high-dimensional probability distribution ?



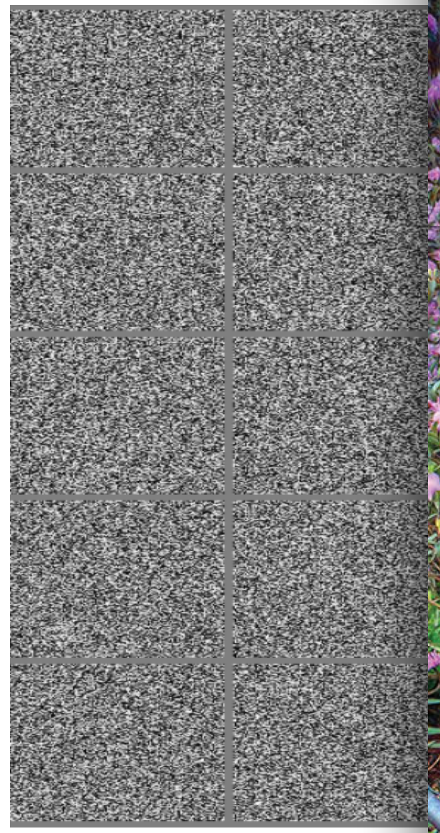


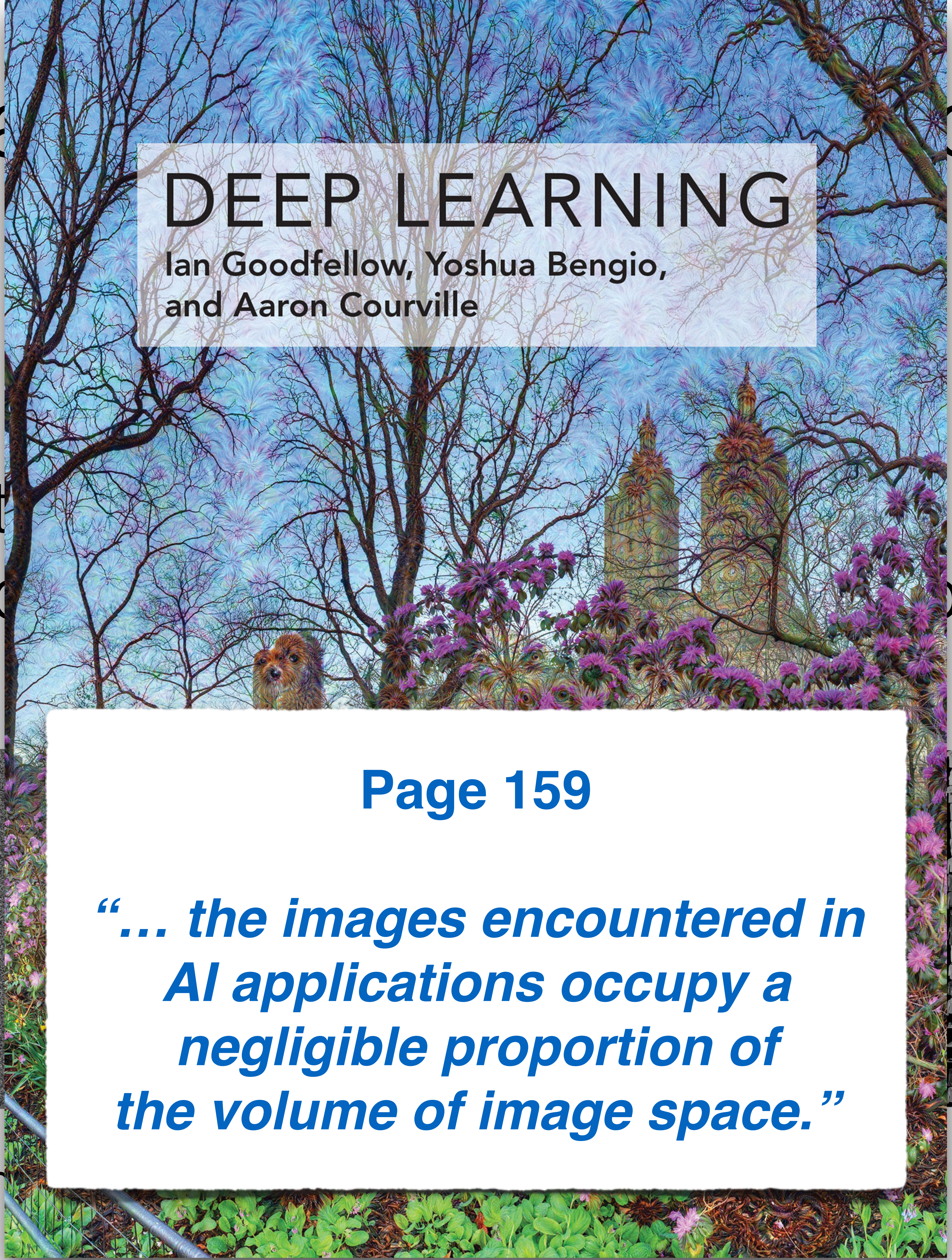
"random" images

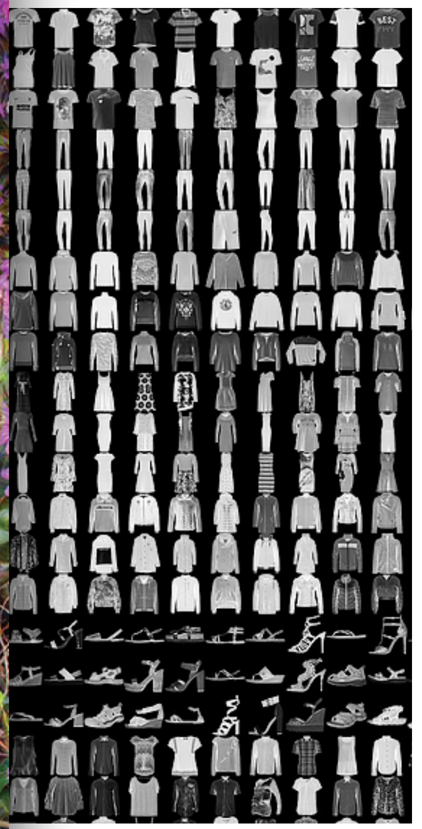
"natural" images

Proba odeling

How t from a
high-d ution ?





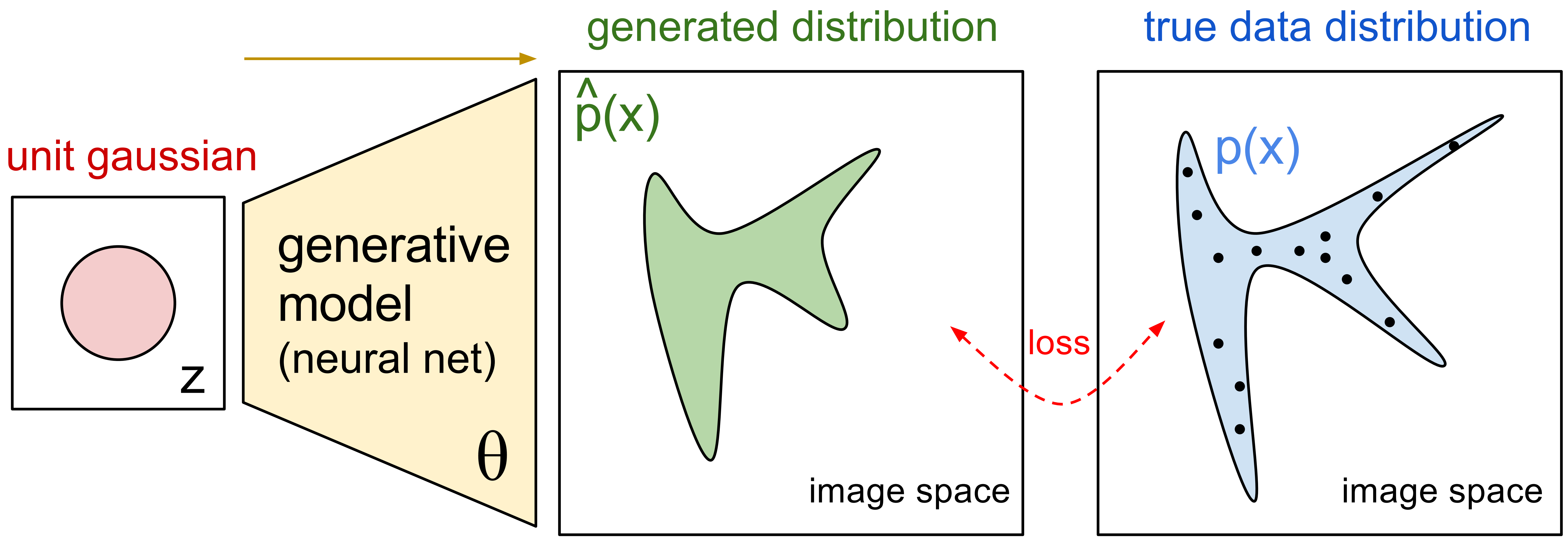


“random

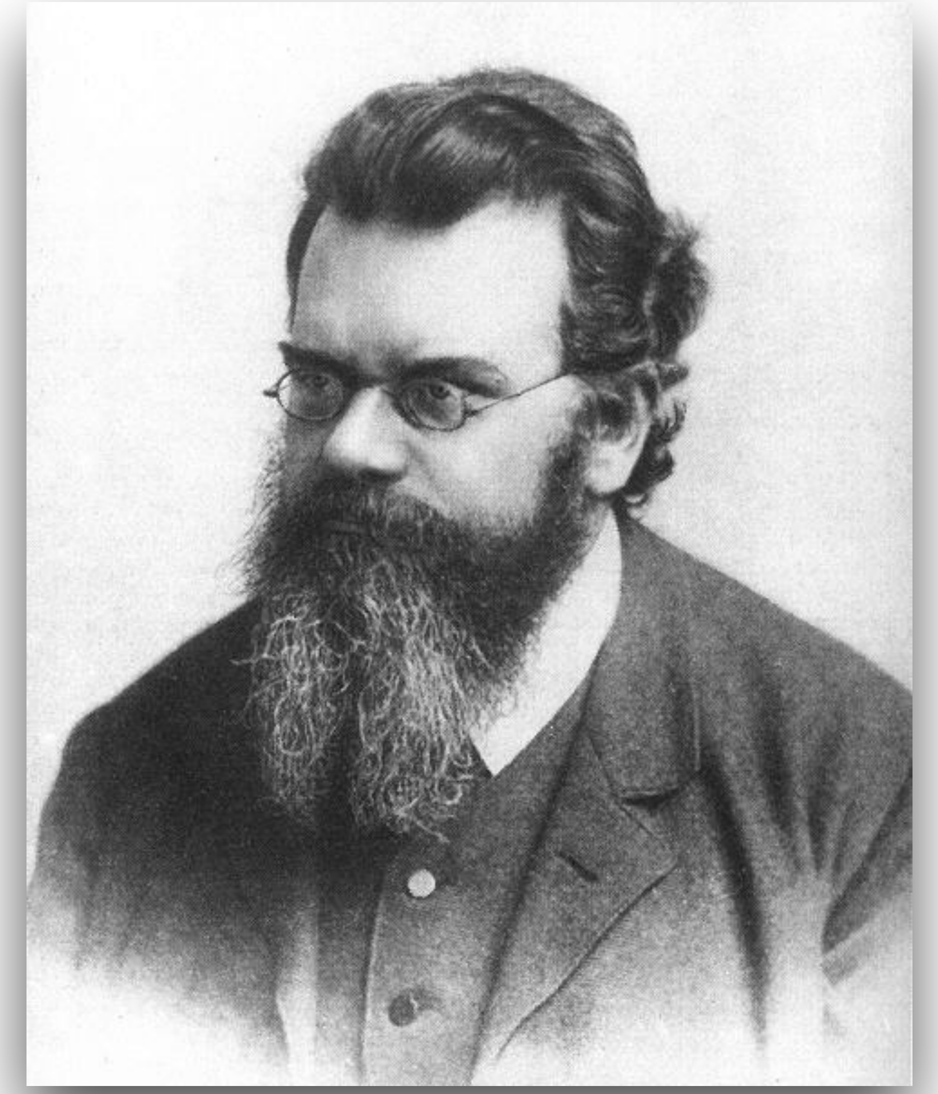
Probabilistic Generative Modeling

$$p(\boldsymbol{x})$$

How to express, learn, and sample from a high-dimensional probability distribution ?



https://blog.openai.com/generative-models/



Boltzmann Machines

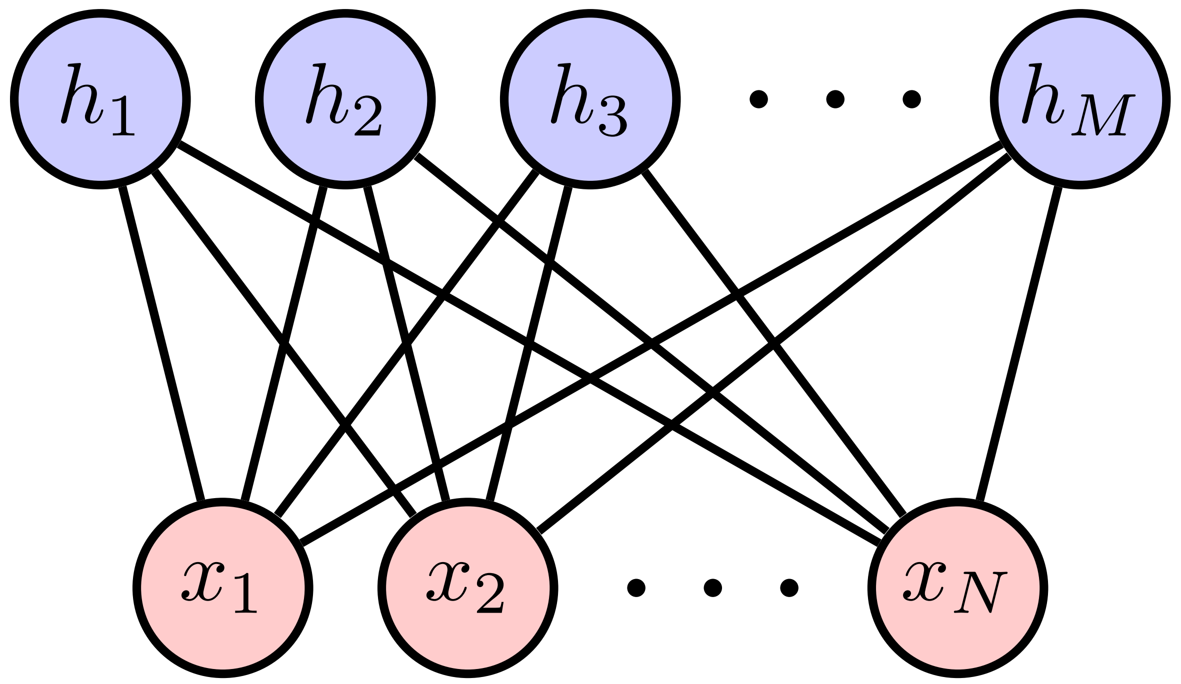
$$p(\boldsymbol{x}) = \frac{e^{-E(\boldsymbol{x})}}{Z}$$

statistical physics

Generative Modeling using Boltzmann Machines

$$\mathcal{L} = -\frac{1}{|\mathcal{T}|}\sum_{\boldsymbol{x}\in\mathcal{T}} \ln p(\boldsymbol{x})$$ Negative log-likelihood loss

Learn



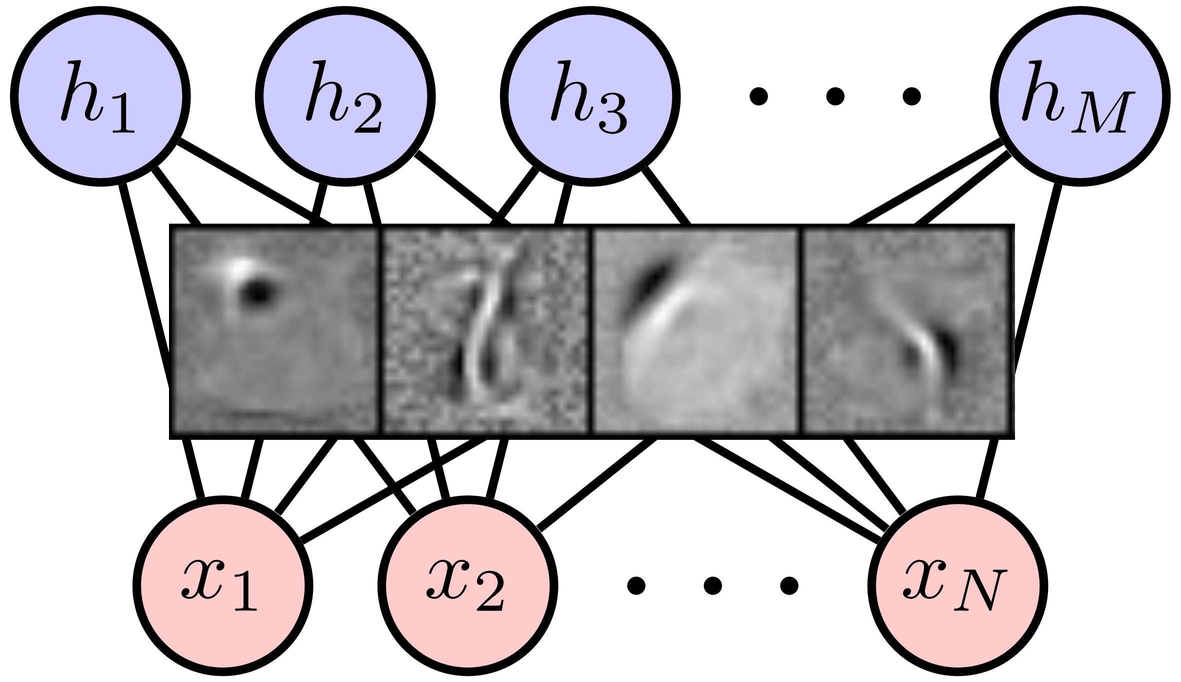
$$\nabla\mathcal{L} = \langle \nabla E \rangle_{\mathrm{data}} - \langle \nabla E \rangle_{\mathrm{model}}$$

Wavefunctions ansatz, Carleo & Troyer, …
Quantum error decoder, Torlai & Melko, …
Quantum state tomography, Torlai et al, next talk

Generative Modeling using Boltzmann Machines

$$\mathcal{L} = -\frac{1}{|\mathcal{T}|}\sum_{\boldsymbol{x}\in\mathcal{T}} \ln p(\boldsymbol{x})$$ Negative log-likelihood loss

Learn



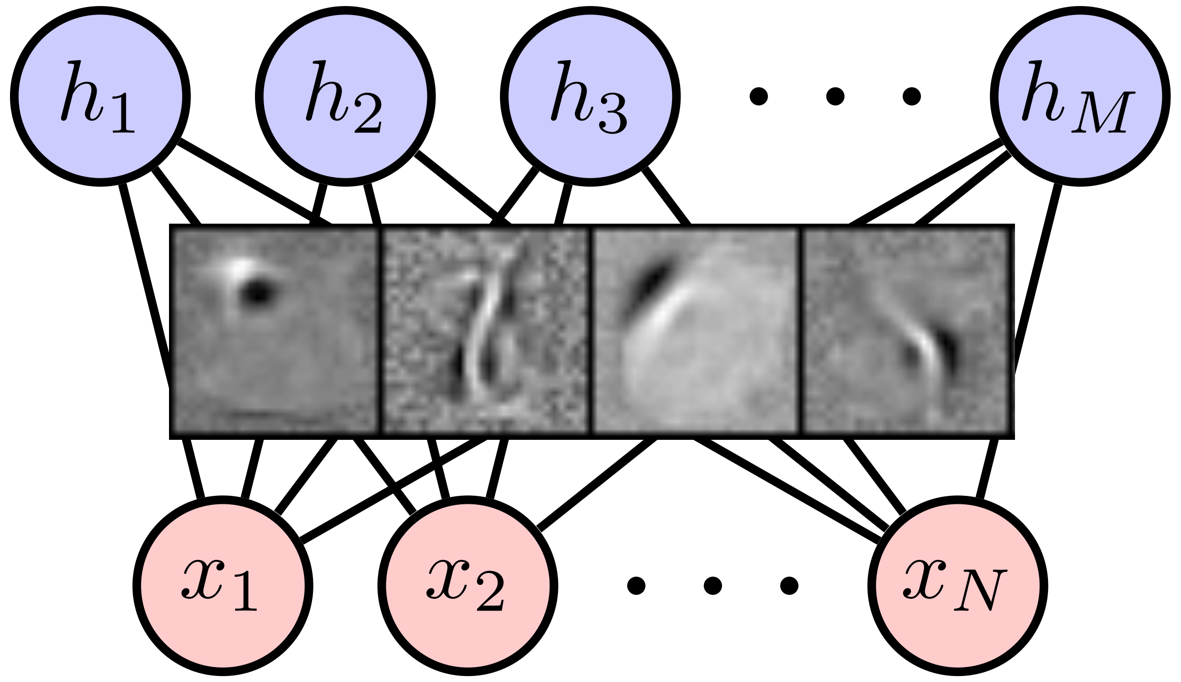
$$\nabla\mathcal{L} = \langle \nabla E\rangle_{\text{data}} - \langle \nabla E\rangle_{\text{model}}$$

Wavefunctions ansatz, Carleo & Troyer, …
Quantum error decoder, Torlai & Melko, …
Quantum state tomography, Torlai et al, next talk

Generative Modeling using Boltzmann Machines

$$\mathcal{L} = -\frac{1}{|\mathcal{T}|}\sum_{\boldsymbol{x}\in\mathcal{T}} \ln p\left(\boldsymbol{x}\right)$$ Negative log-likelihood loss

Learn



Generate

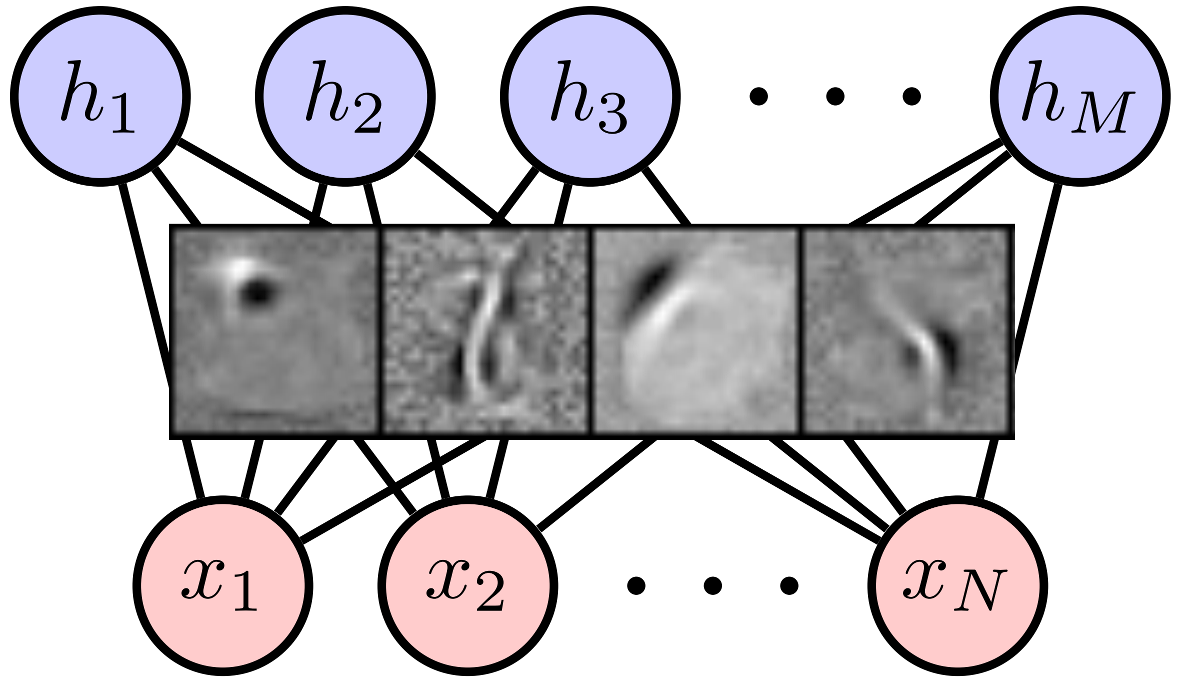
$$\nabla\mathcal{L} = \langle\nabla E\rangle_{\text{data}} - \langle\nabla E\rangle_{\text{model}}$$

Wavefunctions ansatz, Carleo & Troyer, …
Quantum error decoder, Torlai & Melko, …
Quantum state tomography, Torlai et al, next talk

Generative Modeling using Boltzmann Machines

$$\mathcal{L} = -\frac{1}{|\mathcal{T}|}\sum_{\boldsymbol{x}\in\mathcal{T}} \ln p(\boldsymbol{x})$$ Negative log-likelihood loss

Learn

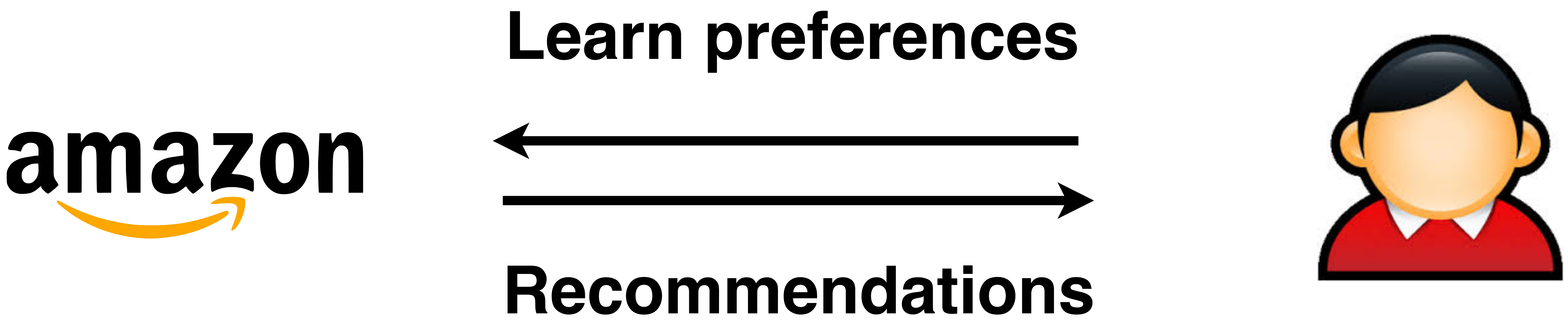


Generate

$$\nabla\mathcal{L} = \langle \nabla E\rangle_{\mathrm{data}} - \langle \nabla E\rangle_{\mathrm{model}}$$

Wavefunctions ansatz, Carleo & Troyer, …
Quantum error decoder, Torlai & Melko, …
Quantum state tomography, Torlai et al, next talk

Monte Carlo update proposals using Boltzmann Machines

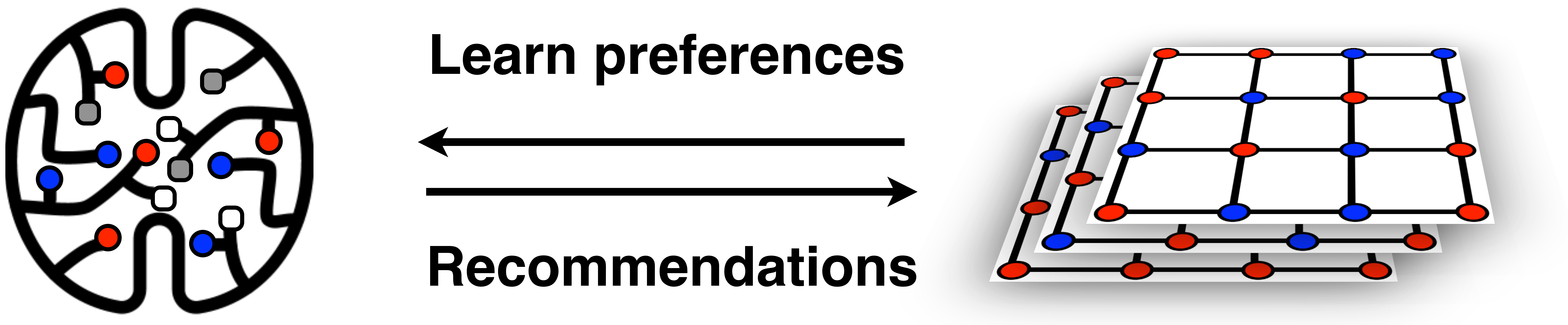


- Use Boltzmann Machines as recommender systems for Monte Carlo simulation of physical problems

Li Huang and LW, 1610.02746

Liu, Qi, Meng, Fu, 1610.03137

Monte Carlo update proposals using Boltzmann Machines

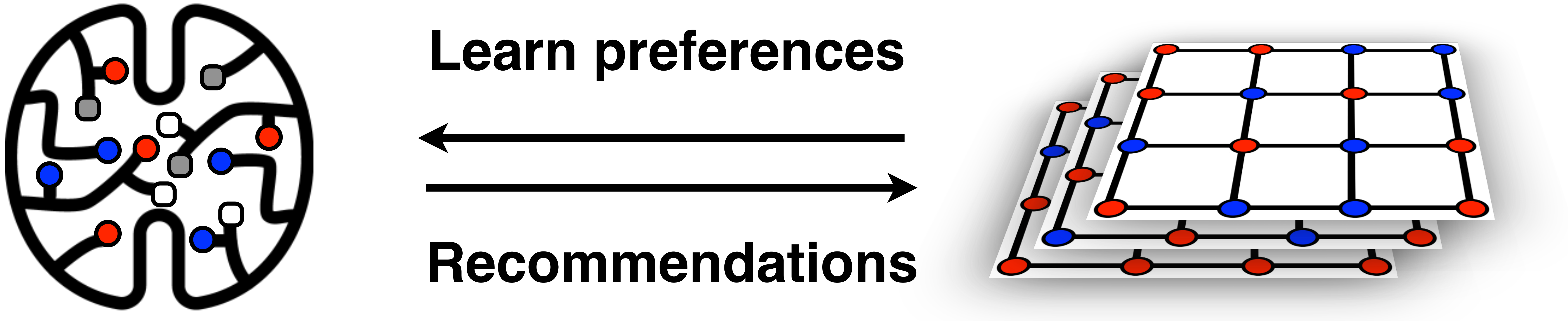


- Use Boltzmann Machines as recommender systems for Monte Carlo simulation of physical problems

Li Huang and LW, 1610.02746

Liu, Qi, Meng, Fu, 1610.03137

Monte Carlo update proposals using Boltzmann Machines



- Use Boltzmann Machines as recommender systems for Monte Carlo simulation of physical problems

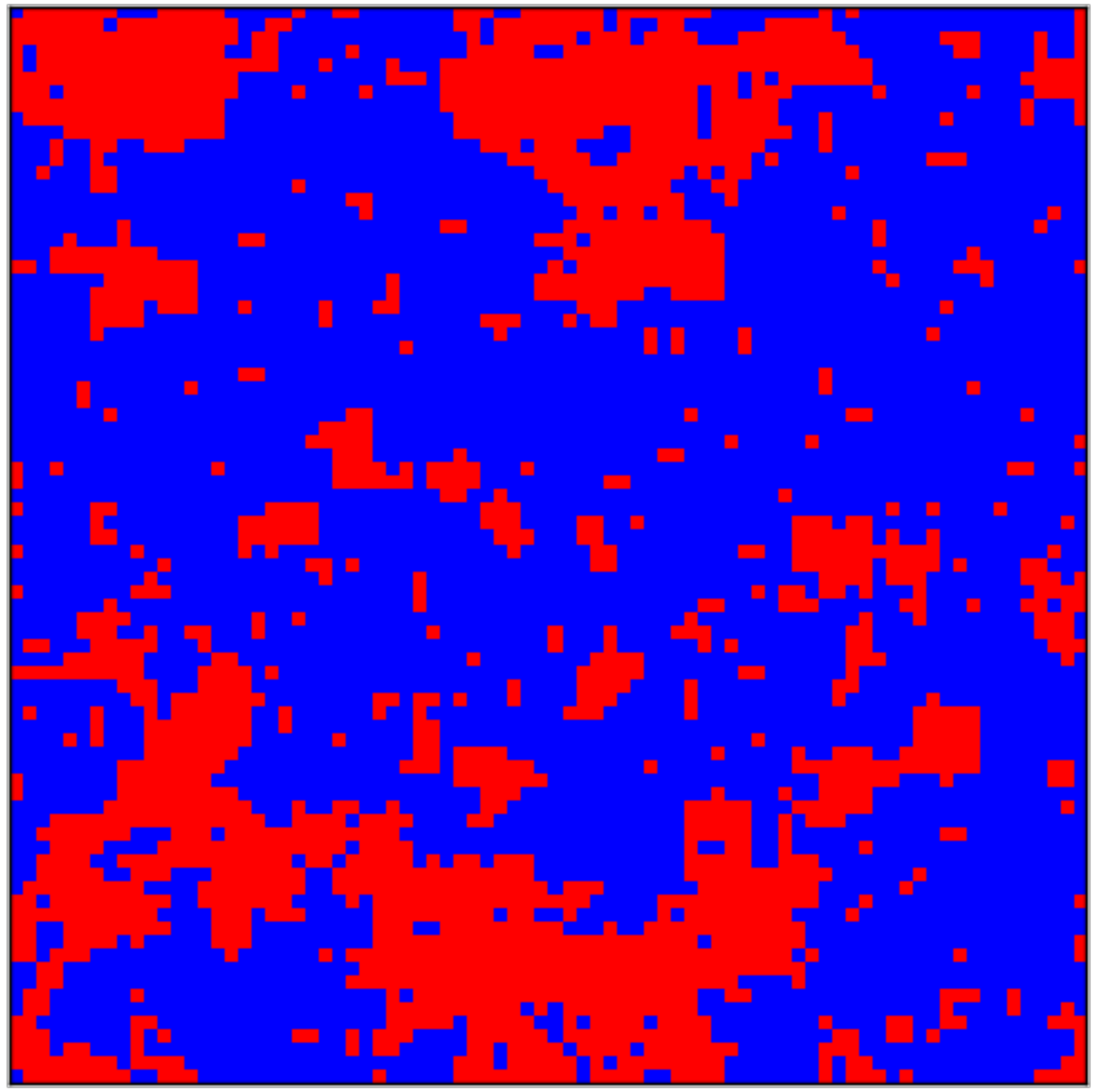
Li Huang and LW, 1610.02746

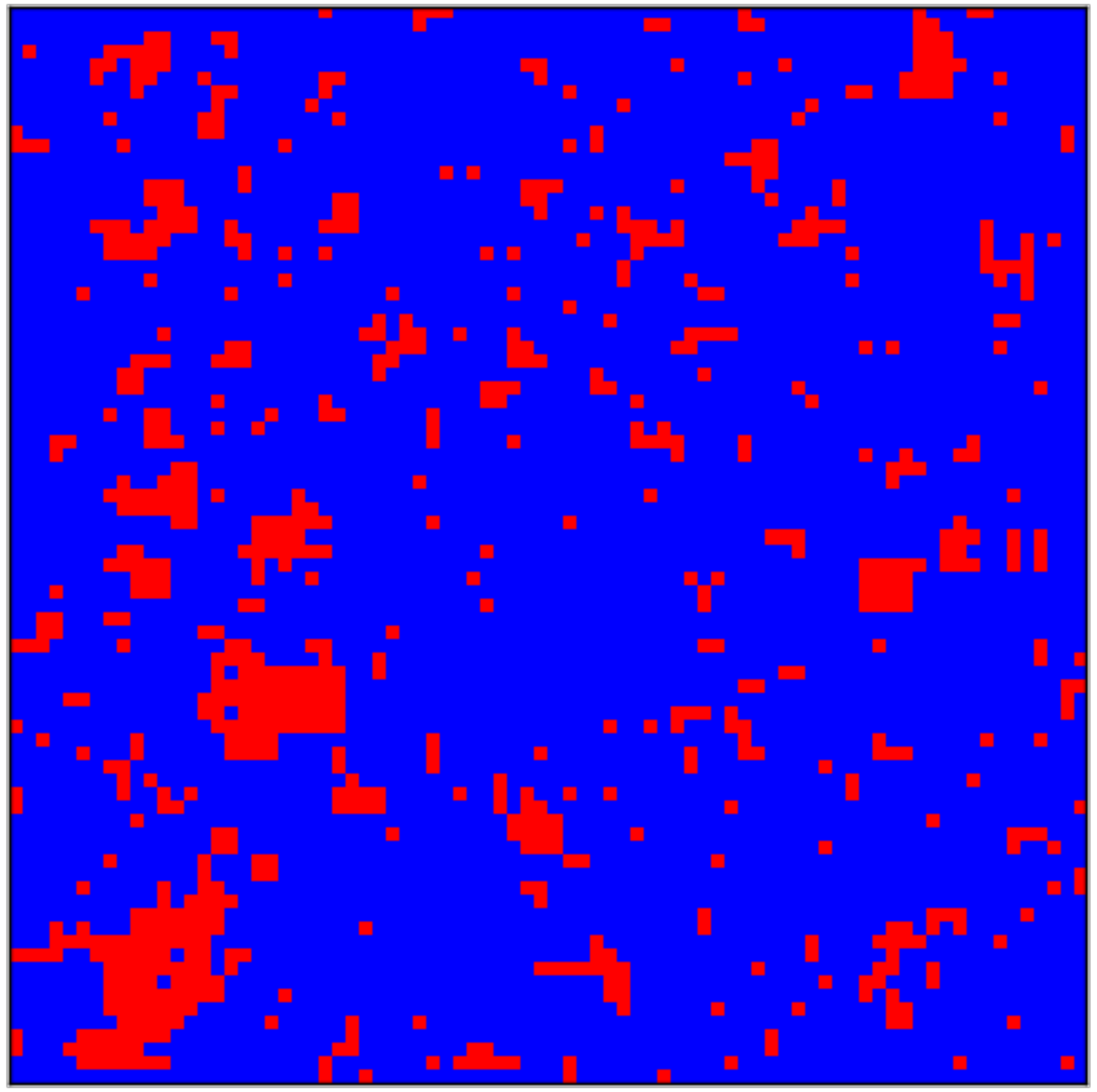
Liu, Qi, Meng, Fu, 1610.03137

- Moreover, BM parametrizes Monte Carlo policies and can explore novel algorithms!

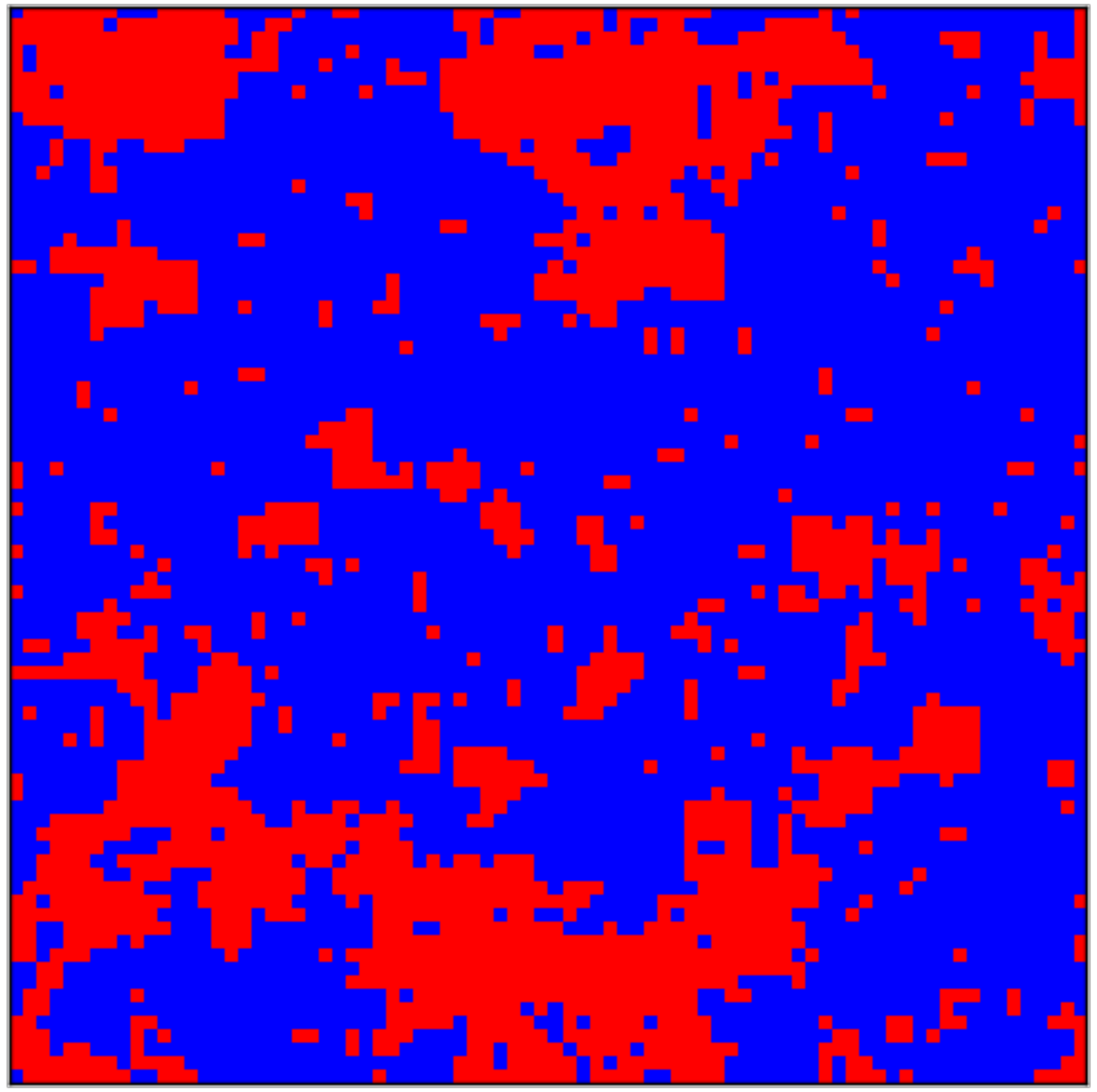
LW, 1702.08586

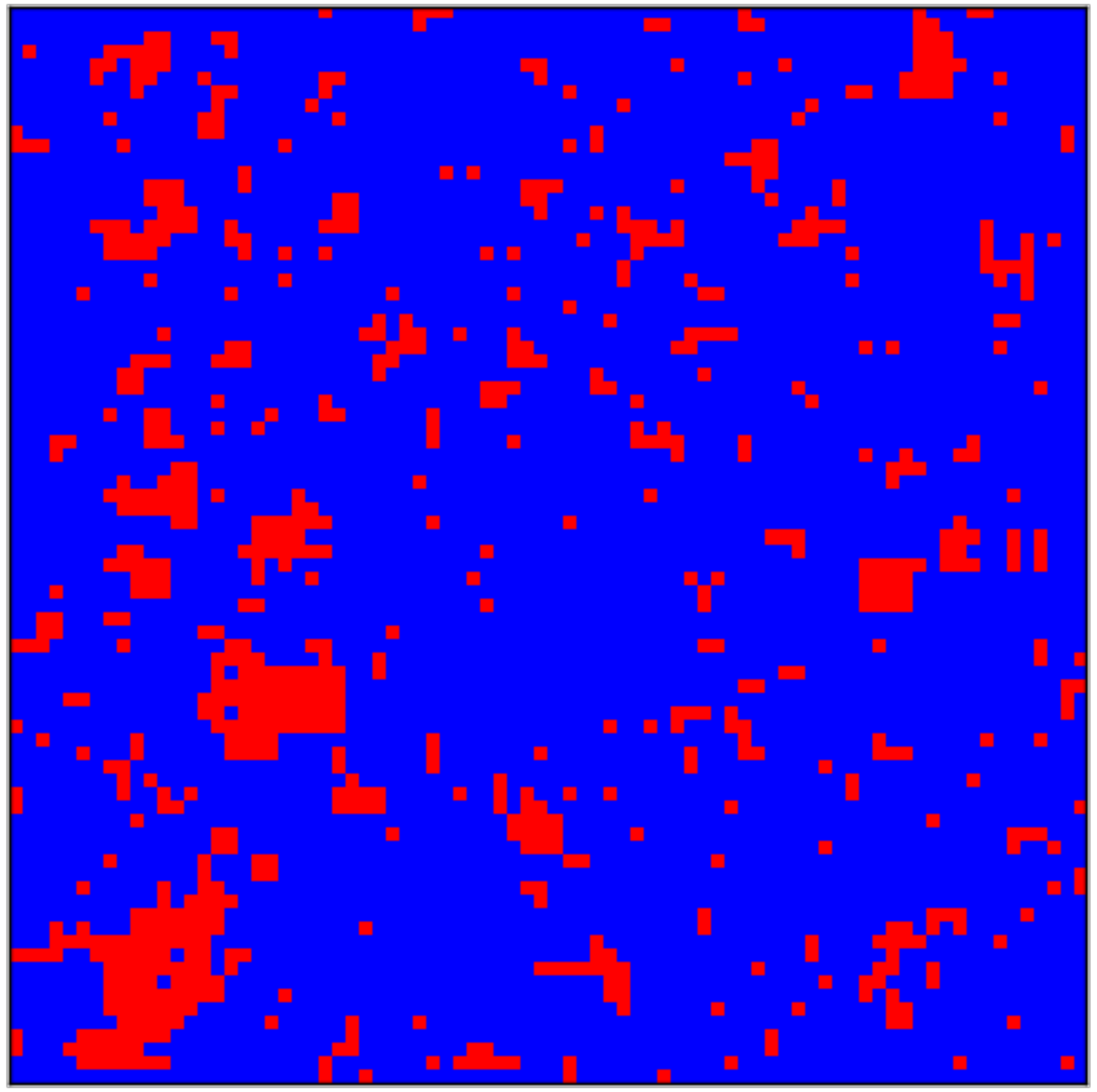
Local vs Cluster update polices



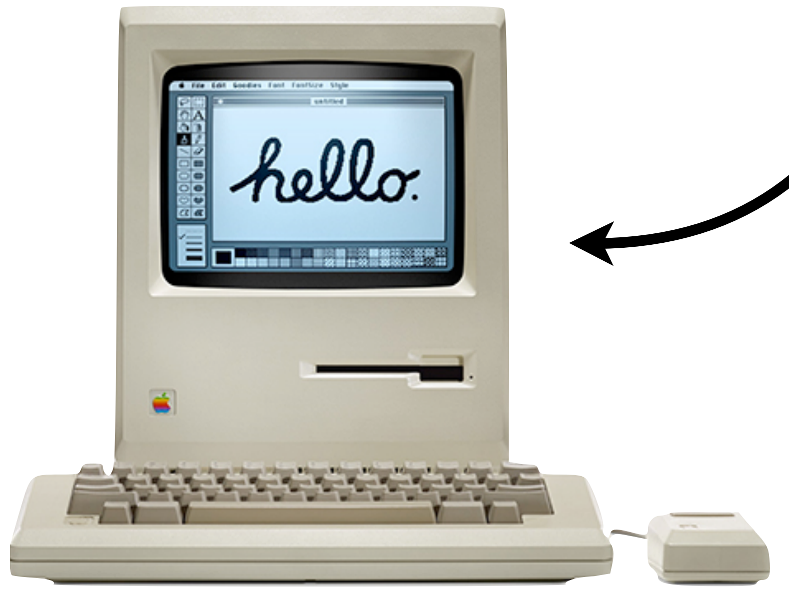


Local vs Cluster update polices

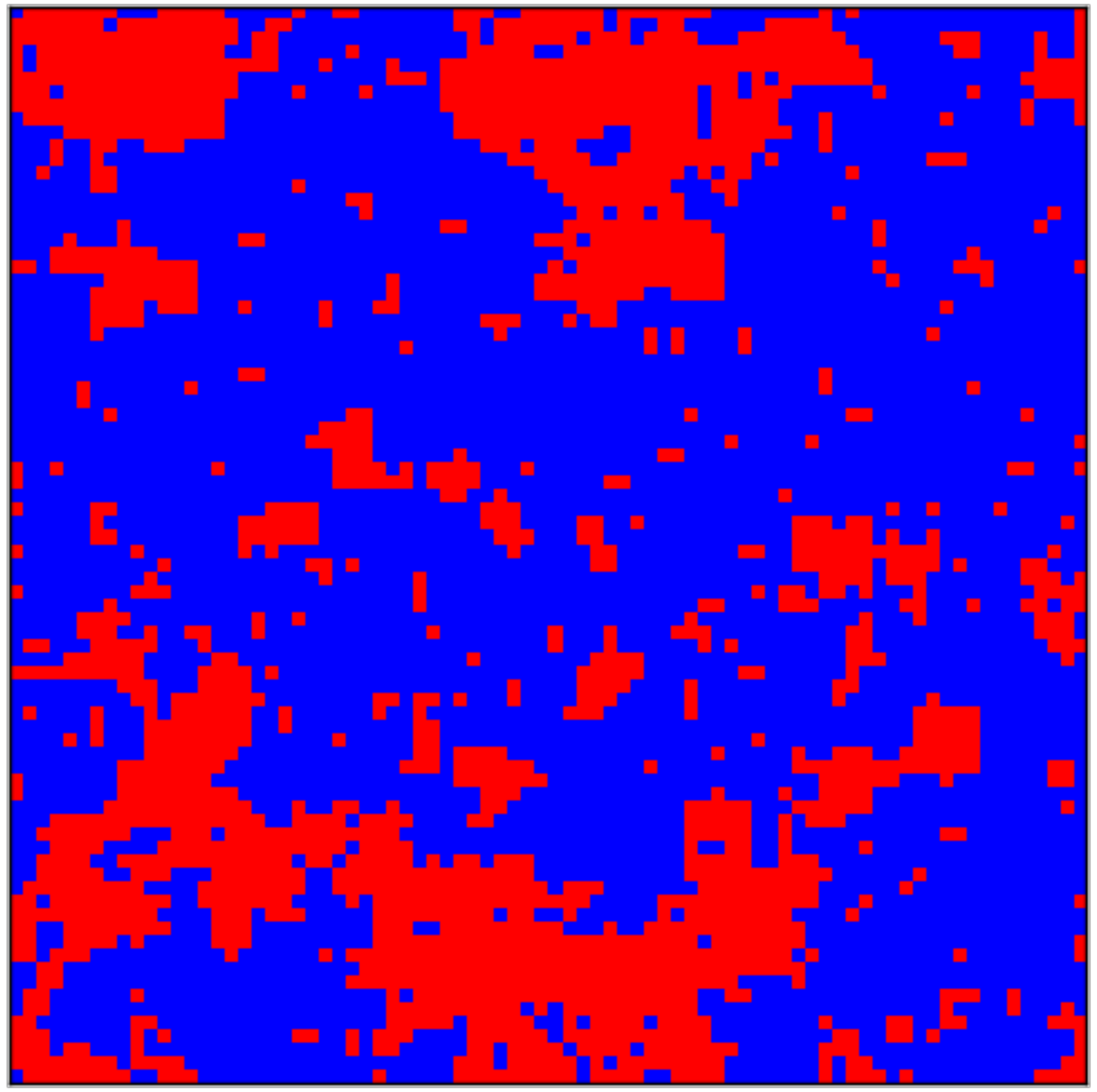


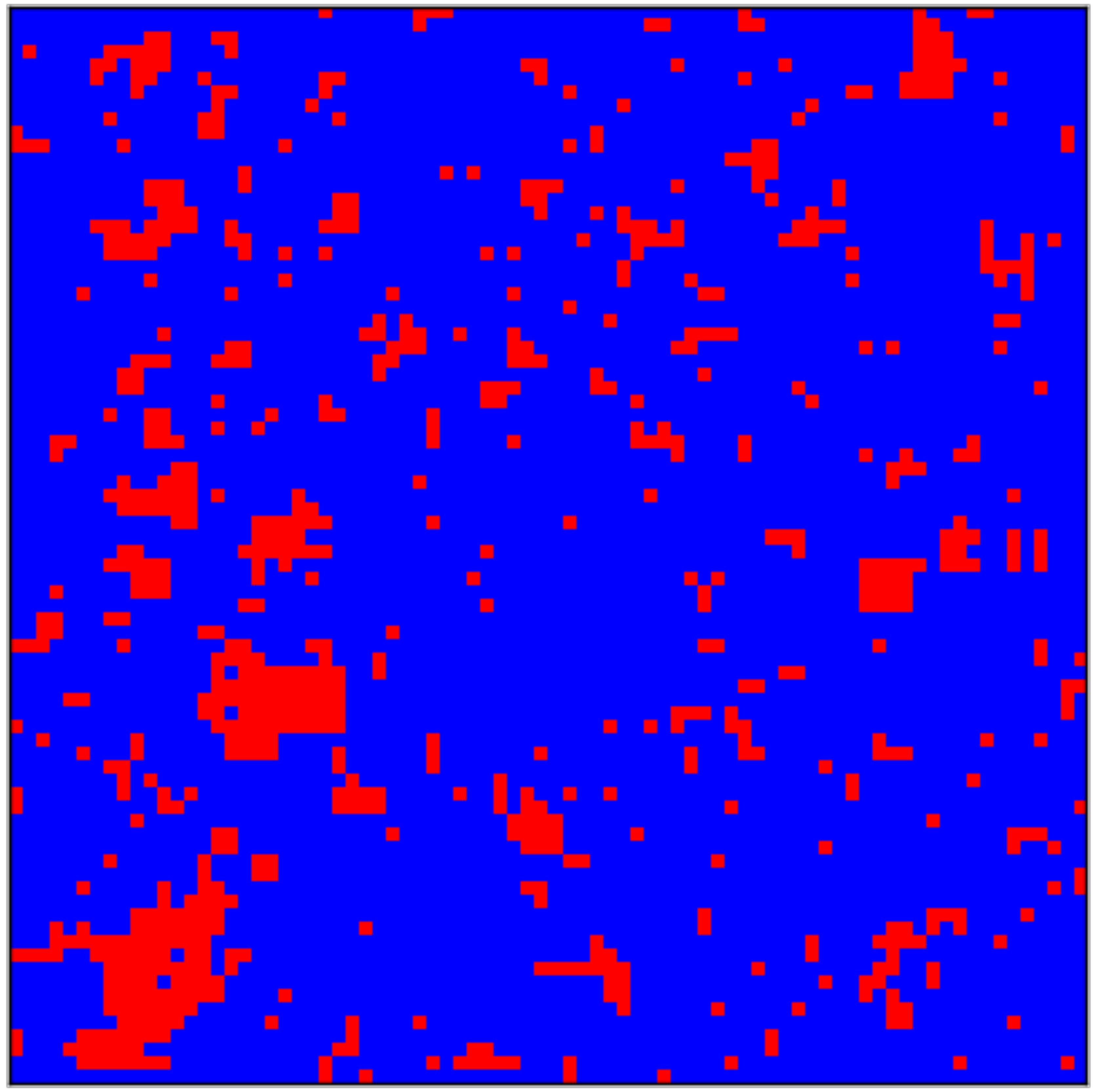


is slower than



Local vs Cluster update polices



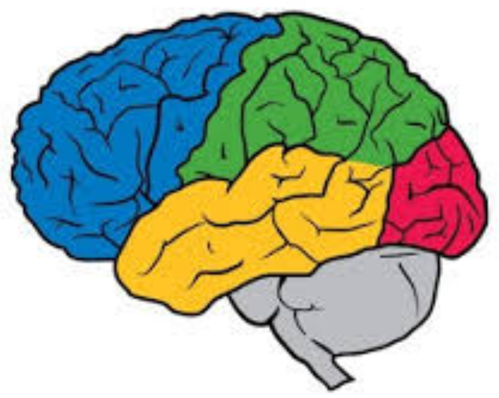


Algorithmic innovation outperforms Moore's law!

Aside: Deep learning Monte Carlo proposals

$$A(\boldsymbol{x} \to \boldsymbol{x}') = \min\left[1, \underbrace{\boxed{\frac{q(\boldsymbol{x}' \to \boldsymbol{x})}{q(\boldsymbol{x} \to \boldsymbol{x}')}}}_{\text{Policy}} \cdot \underbrace{\boxed{\frac{\pi(\boldsymbol{x}')}{\pi(\boldsymbol{x})}}}_{\text{Physics}}\right]$$

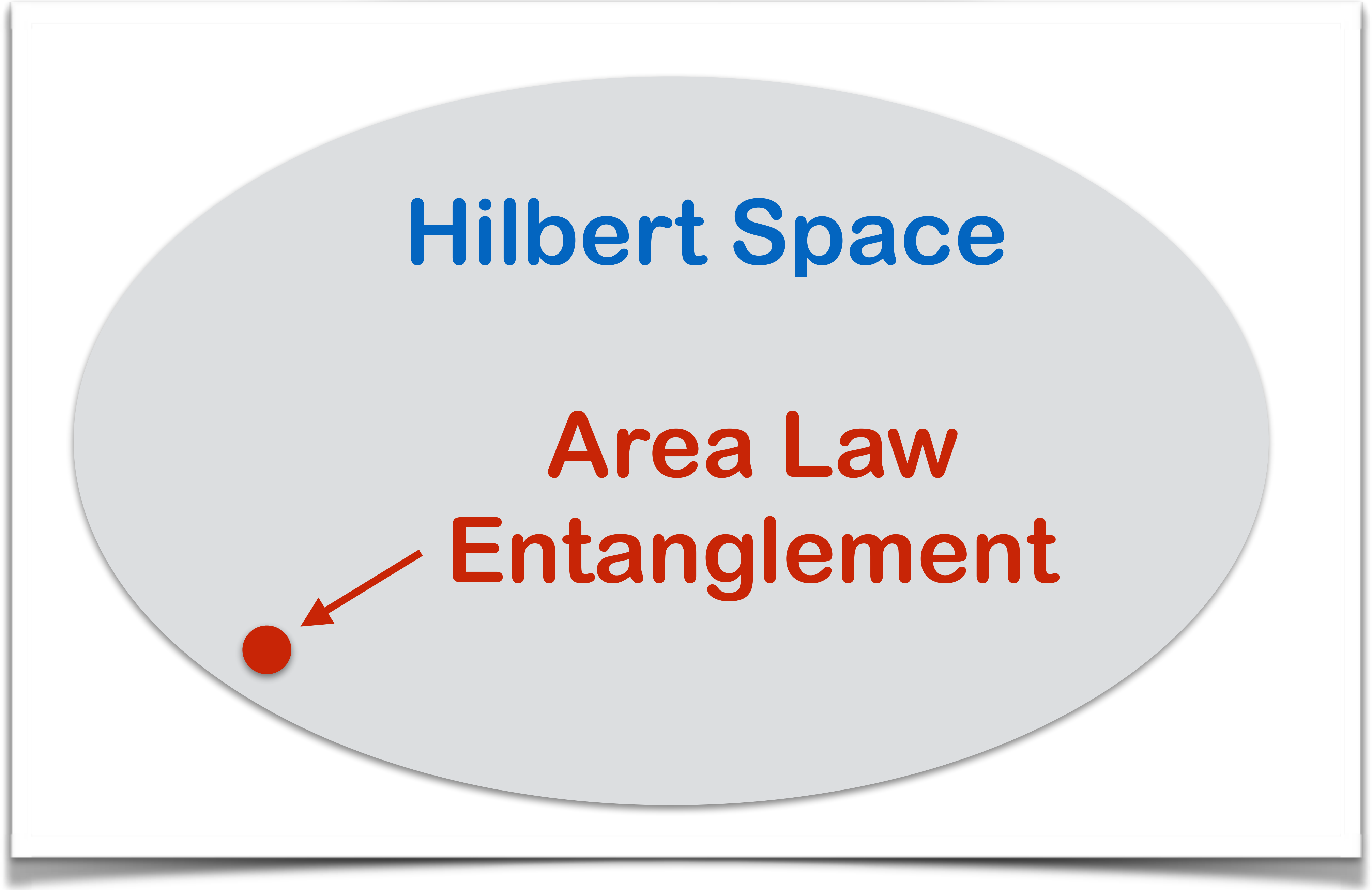
- Adversarial training for MCMC 1706.07561
- Generalize hybrid MC using neural networks 1711.09268
- Probabilistic programs as proposals 1801.03612
- Neural Network Renormalization Group 1802.02840



Born Machines

$$p(\boldsymbol{x}) = \frac{|\Psi(\boldsymbol{x})|^2}{\mathcal{Z}}$$

quantum physics

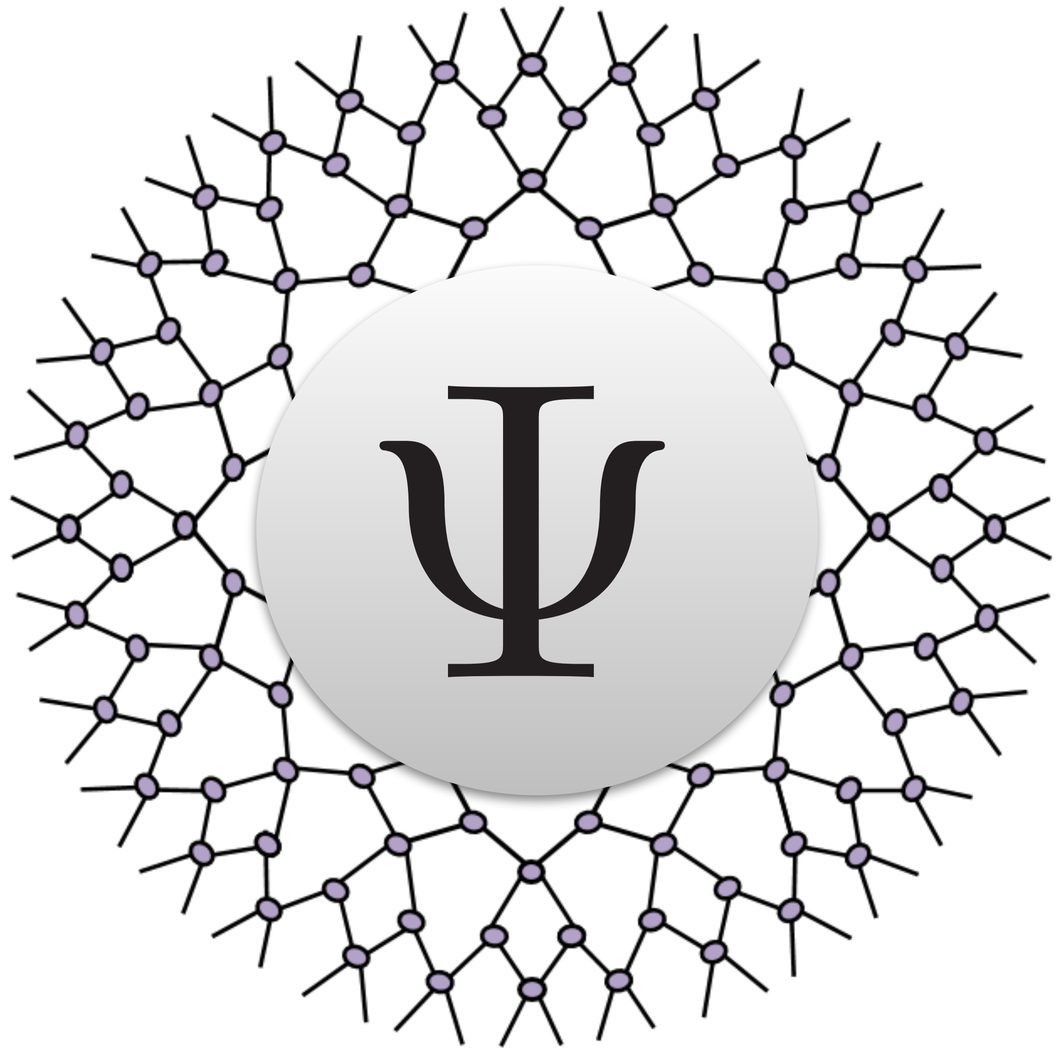


Born Machines

$$p(\boldsymbol{x}) = \frac{|\Psi(\boldsymbol{x})|^2}{\mathcal{Z}}$$

quantum physics

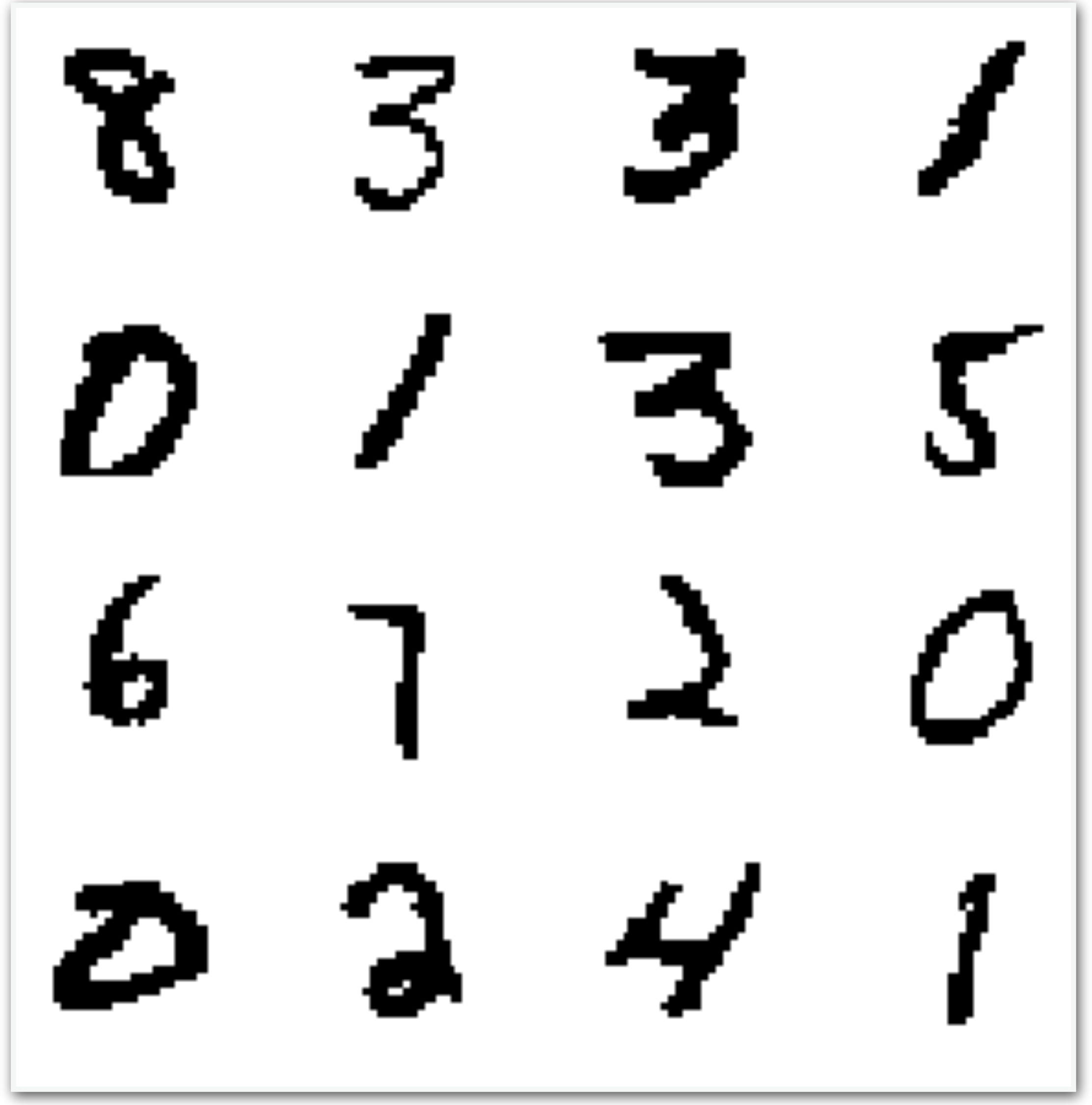
Quantum inspired generative modeling

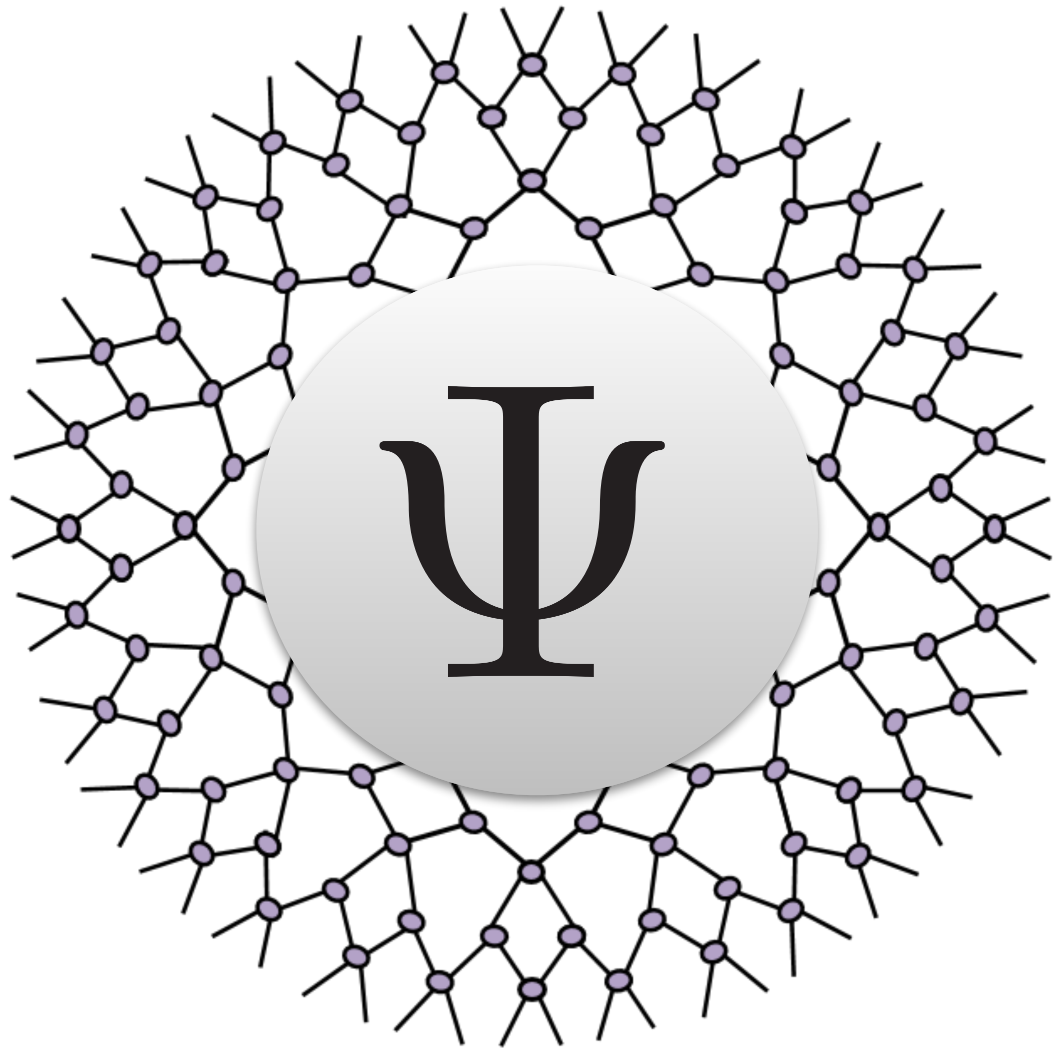


Han, Wang, Fan, LW, Zhang, 1709.01662

Quantum inspired generative modeling

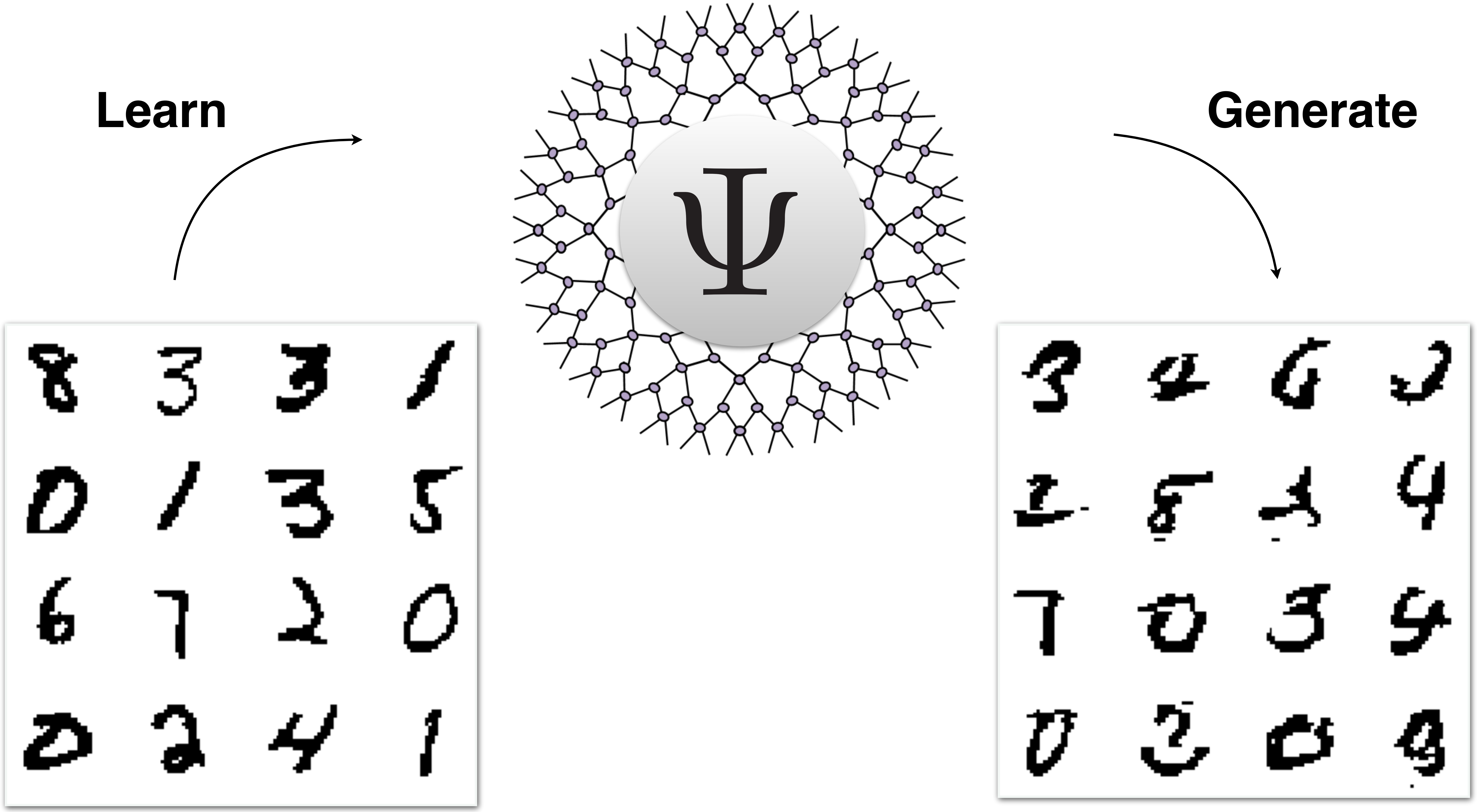
Learn





Han, Wang, Fan, LW, Zhang, 1709.01662

Quantum inspired generative modeling

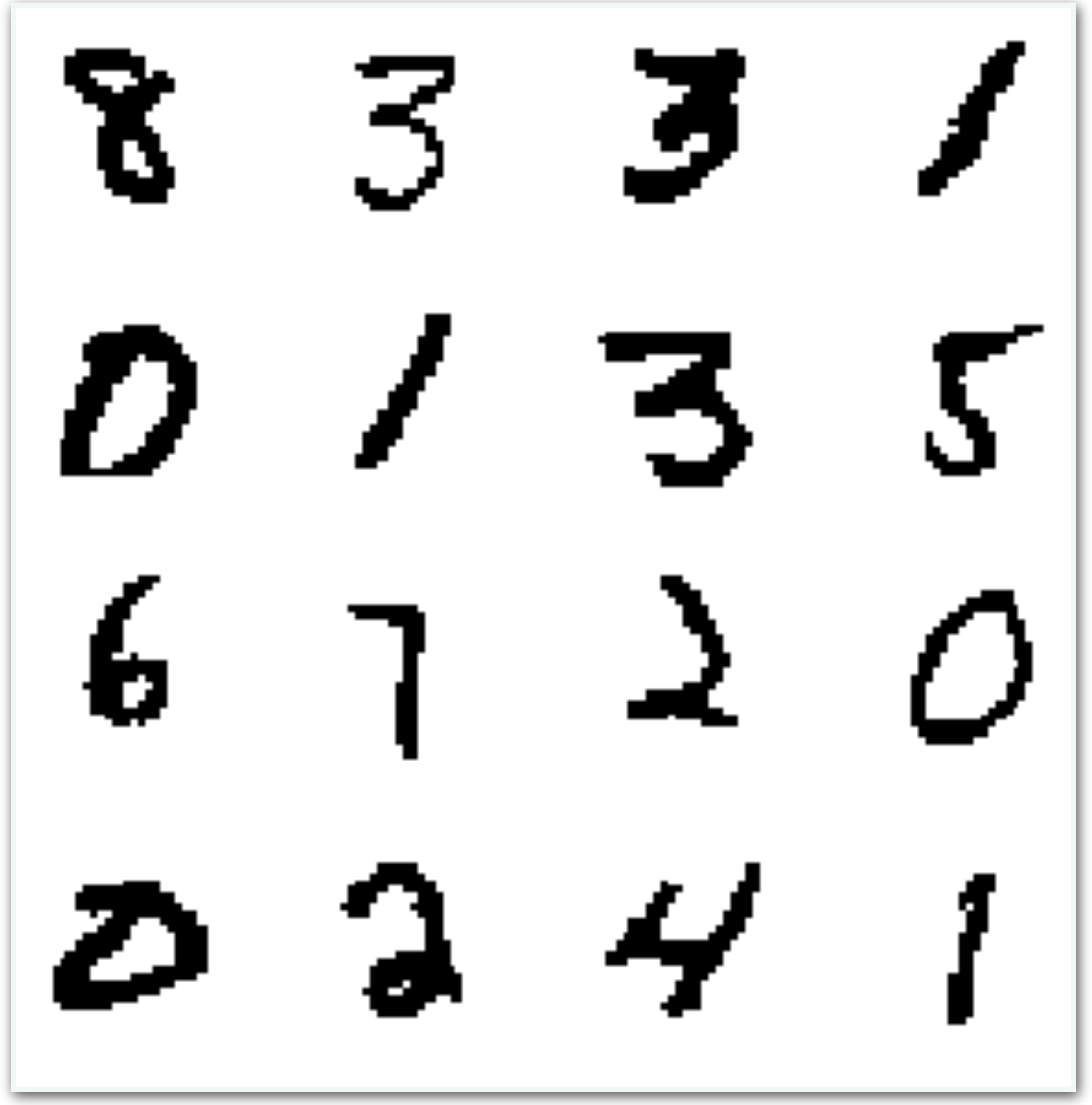


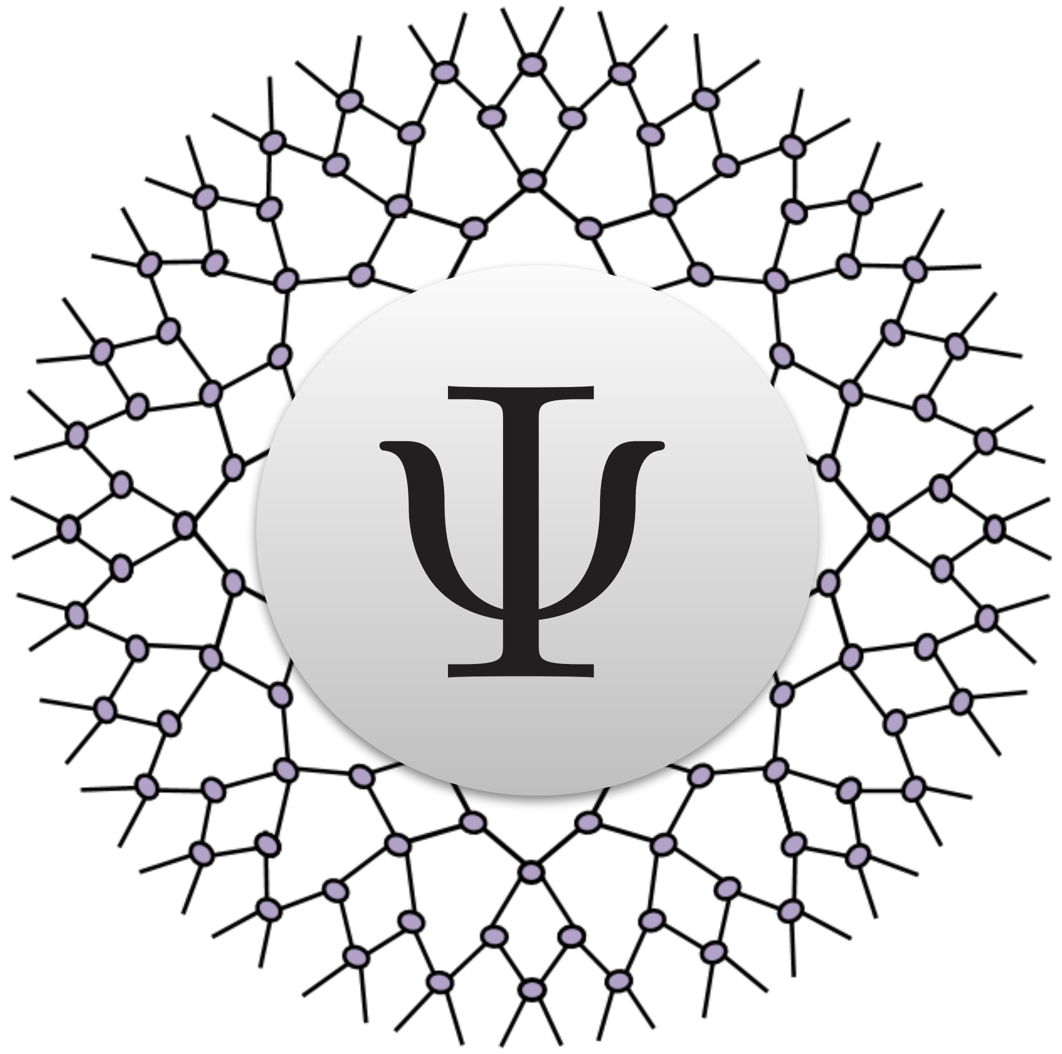
Han, Wang, Fan, LW, Zhang, 1709.01662

Quantum inspired generative modeling

Learn

Generate

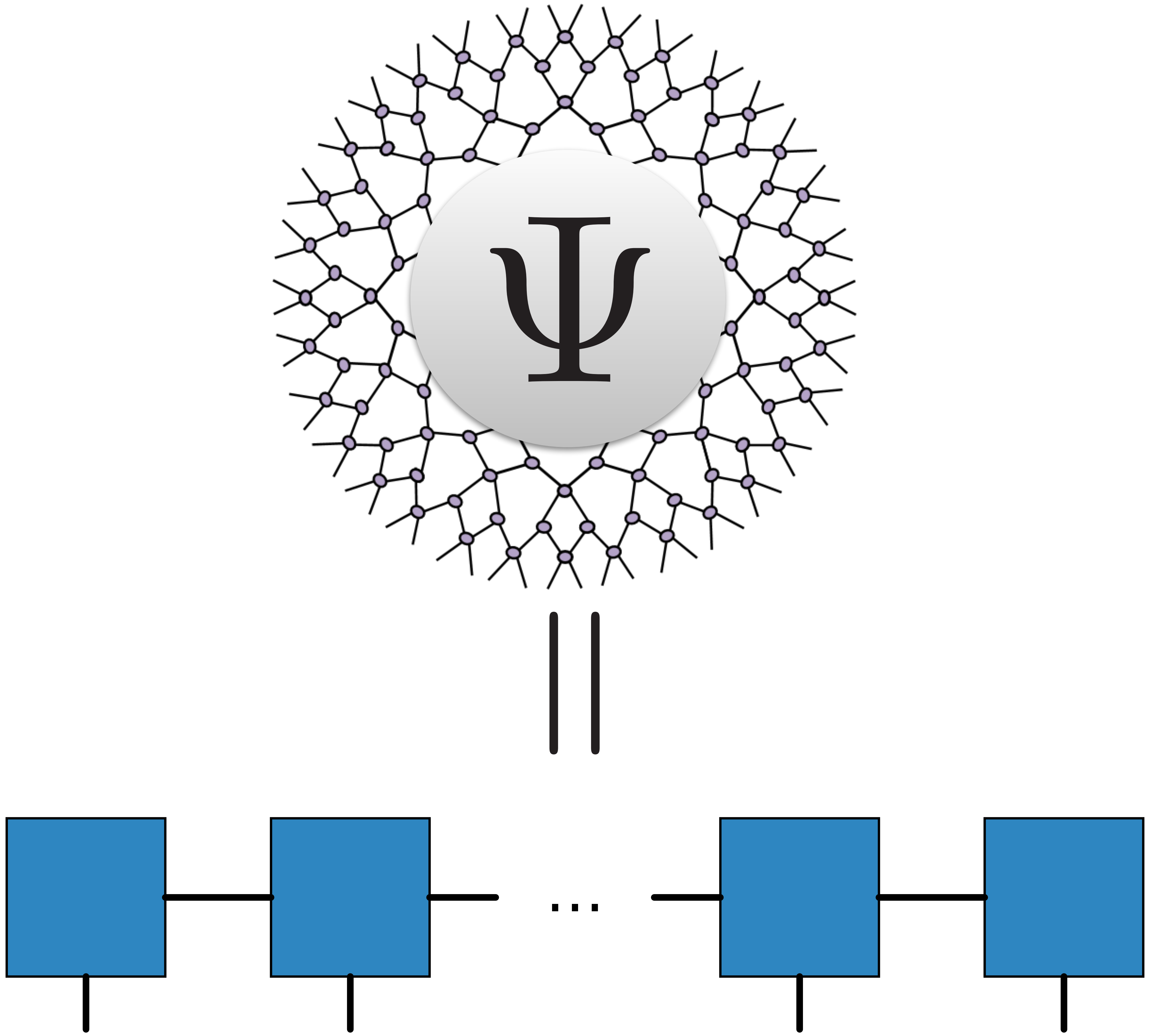




“Teach a quantum state to write digits”

Han, Wang, Fan, LW, Zhang, 1709.01662

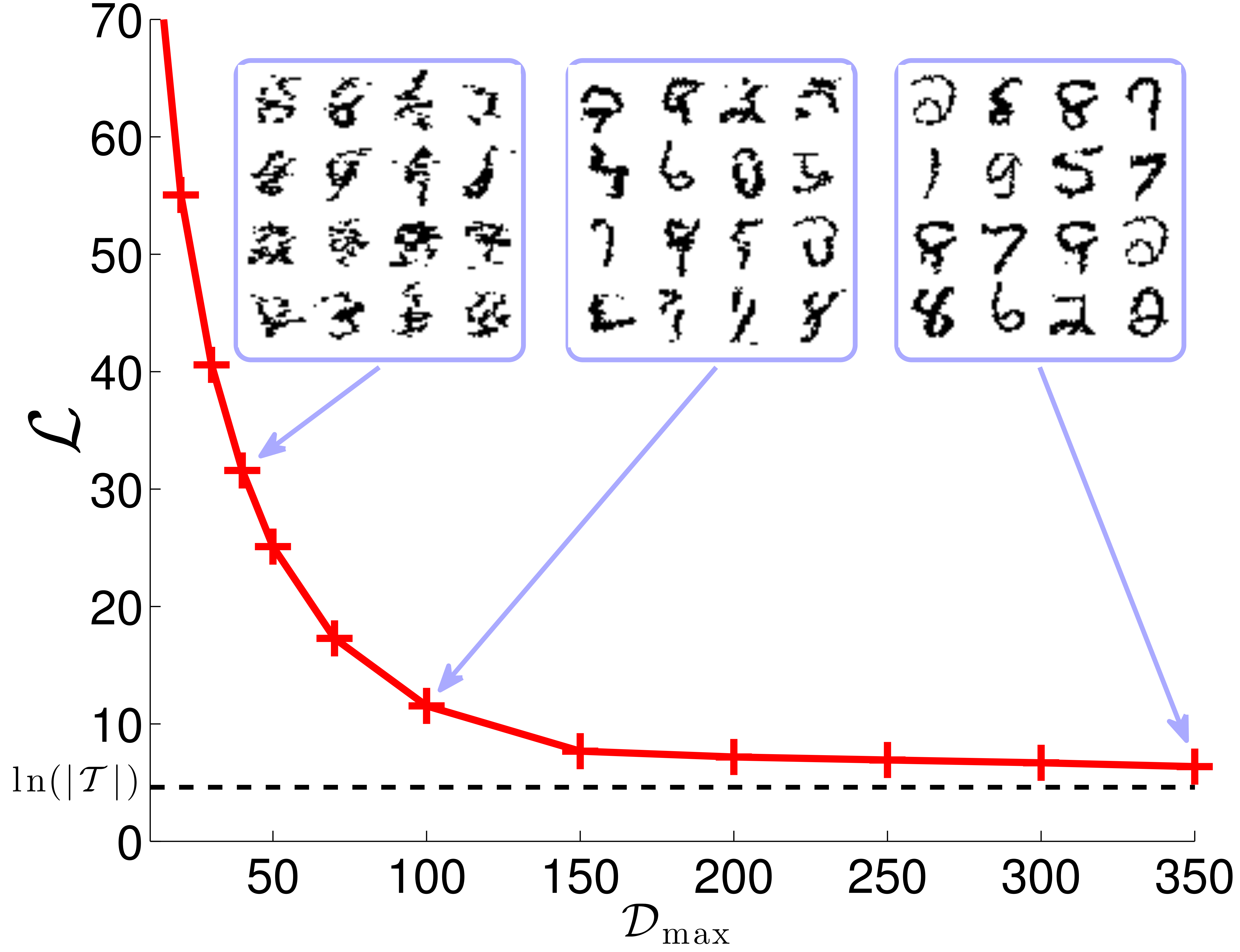
Generative modeling using Tensor Network States



cf. Stoudenmire et al Thursday 8:00AM, Liu et al, 1710.04833

Connections to Boltzmann Machines, Chen, Cheng, Xie, LW, Xiang 1701.04831

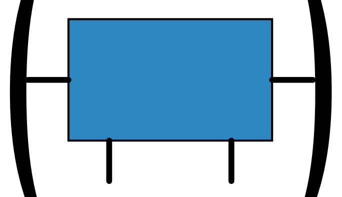
What does it learn ?



Captures longer range correlations with larger bond dimensions

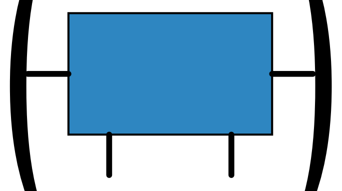
Why bother ?

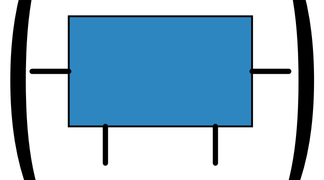
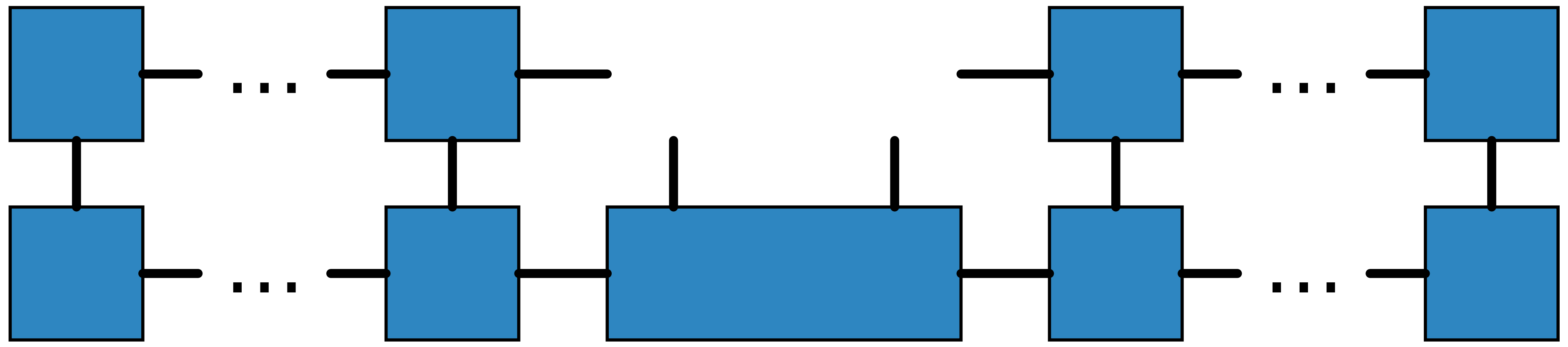
Advantage-I: Unbiased Gradient

$\partial \mathcal{L} \big/ \partial$ () the key thing to compute for optimization

Efficient & Unbiased learning compared to intractable partition function in Boltzmann Machines

Advantage-I: Unbiased Gradient

$\partial \mathcal{L} \Big/ \partial \Big($  $\Big)$ the key thing to compute for optimization

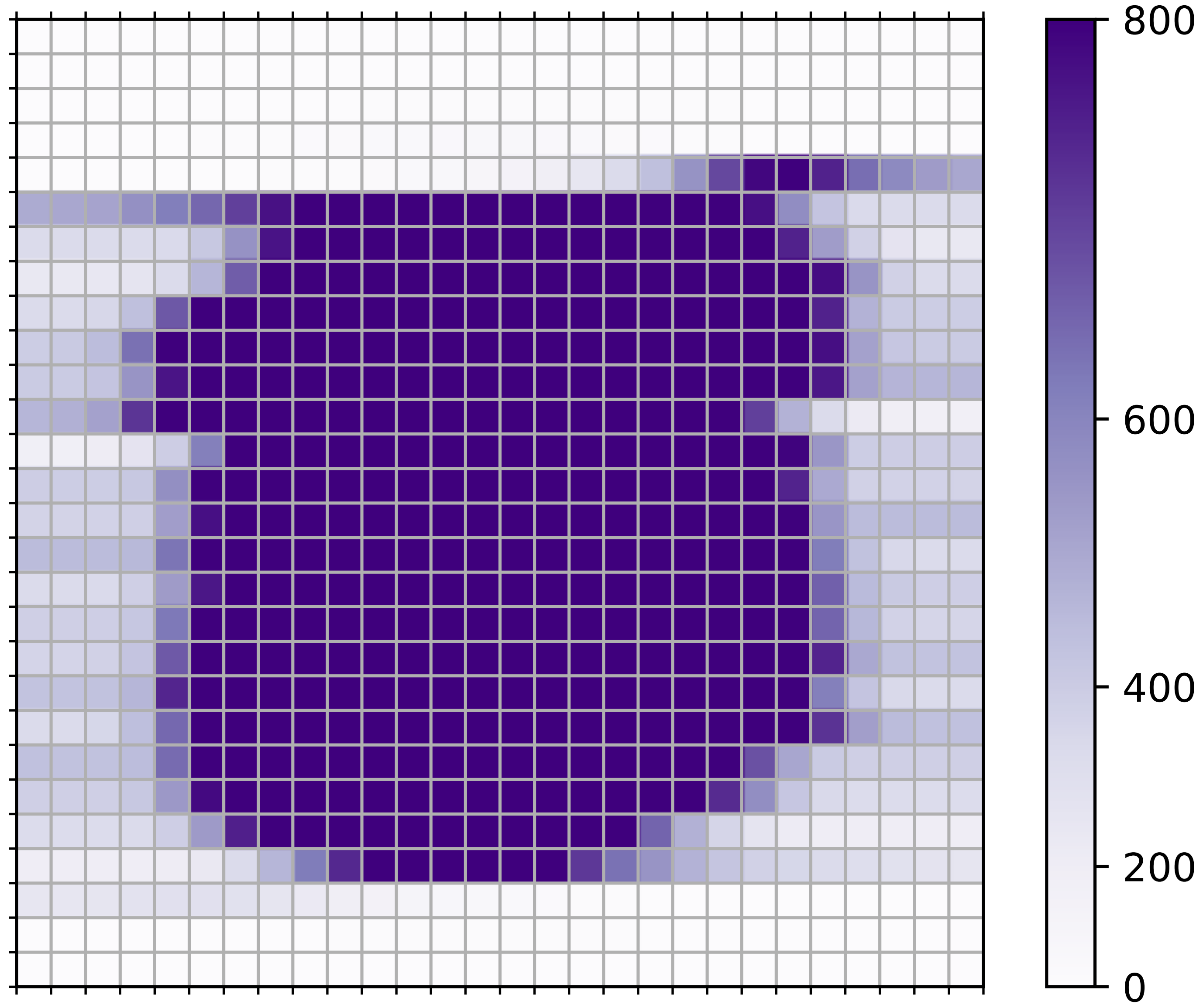
$\partial \mathcal{Z} \Big/ \partial \Big($  $\Big) = 2 \times$ 

Efficient & Unbiased learning compared to
intractable partition function in Boltzmann Machines

Advantage-II: Adaptive Learning

Training images

Bond dimensions



Bond dimensions grow adaptively during learning, thus
dynamically increase expressibility instead of fixing # of params

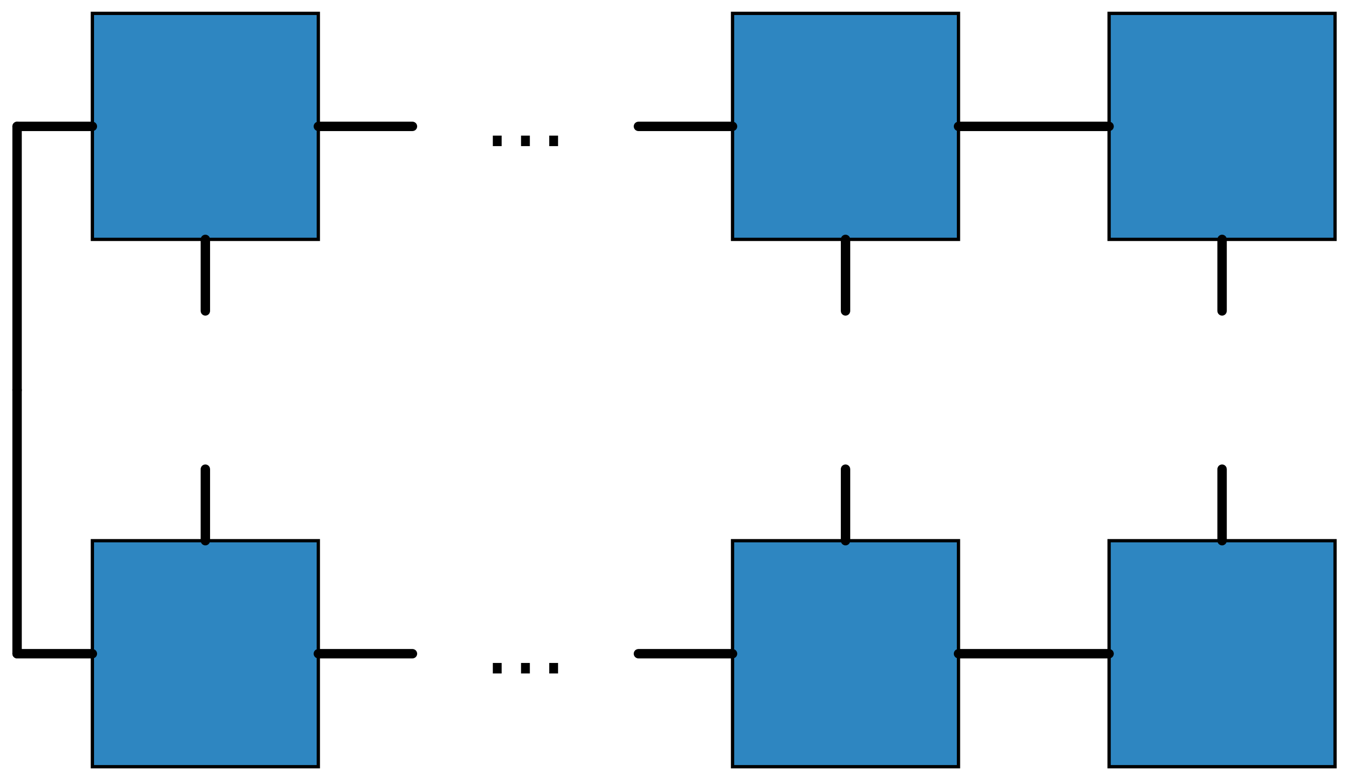
MPS language model:1708.01525, 1711.01416

Advantage-III: Direct Generation

$$p(\boldsymbol{x}) = \prod_i \frac{p(\boldsymbol{x}_{<i+1})}{p(\boldsymbol{x}_{<i})}$$

Ferris & Vidal 2012

$$p(\boldsymbol{x}_{<i}) =$$



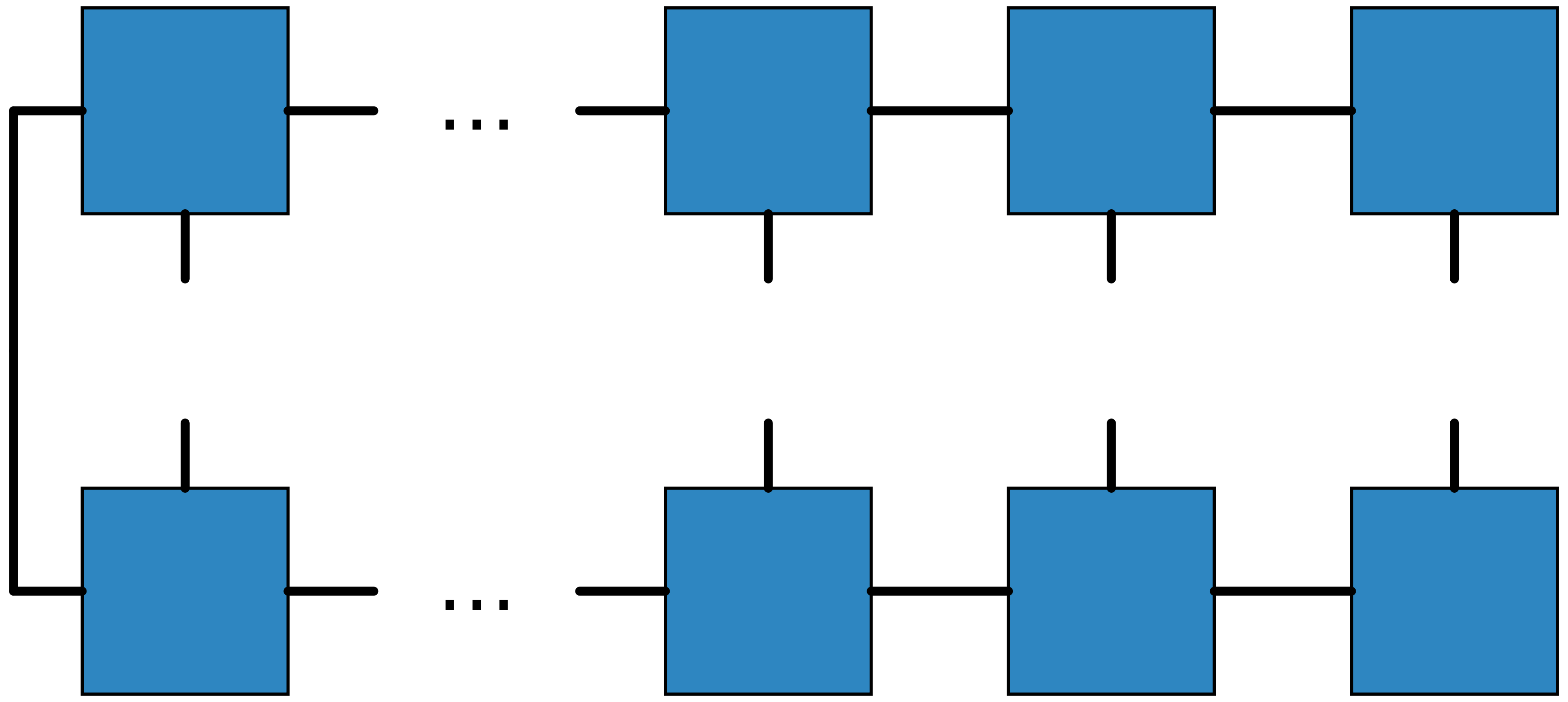
No auto-correlation compared to
slow mixing Gibbs sampling of Boltzmann Machines

Advantage-III: Direct Generation

$$p(\boldsymbol{x}) = \prod_i \frac{p(\boldsymbol{x}_{<i+1})}{p(\boldsymbol{x}_{<i})}$$

Ferris & Vidal 2012

$p(\boldsymbol{x}_{<i+1}) =$



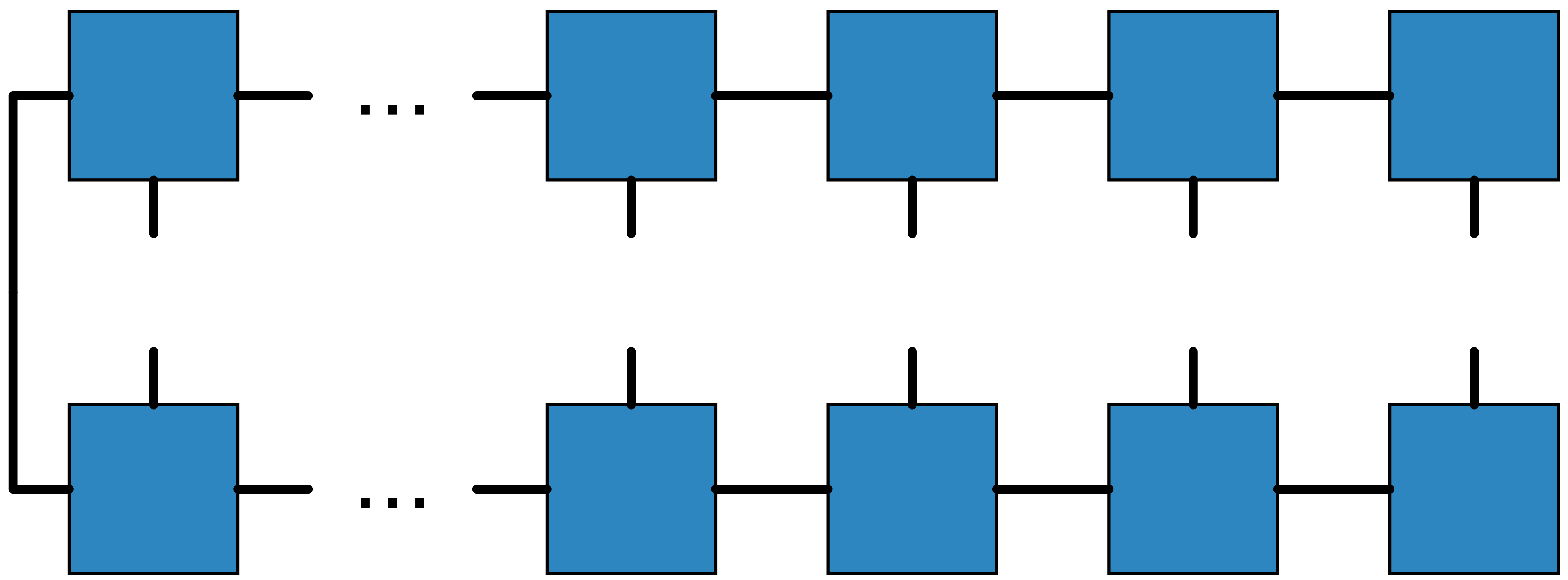
No auto-correlation compared to
slow mixing Gibbs sampling of Boltzmann Machines

Advantage-III: Direct Generation

$$p(\boldsymbol{x}) = \prod_i \frac{p(\boldsymbol{x}_{<i+1})}{p(\boldsymbol{x}_{<i})}$$

Ferris & Vidal 2012

$$p(\boldsymbol{x}_{<i+2}) =$$



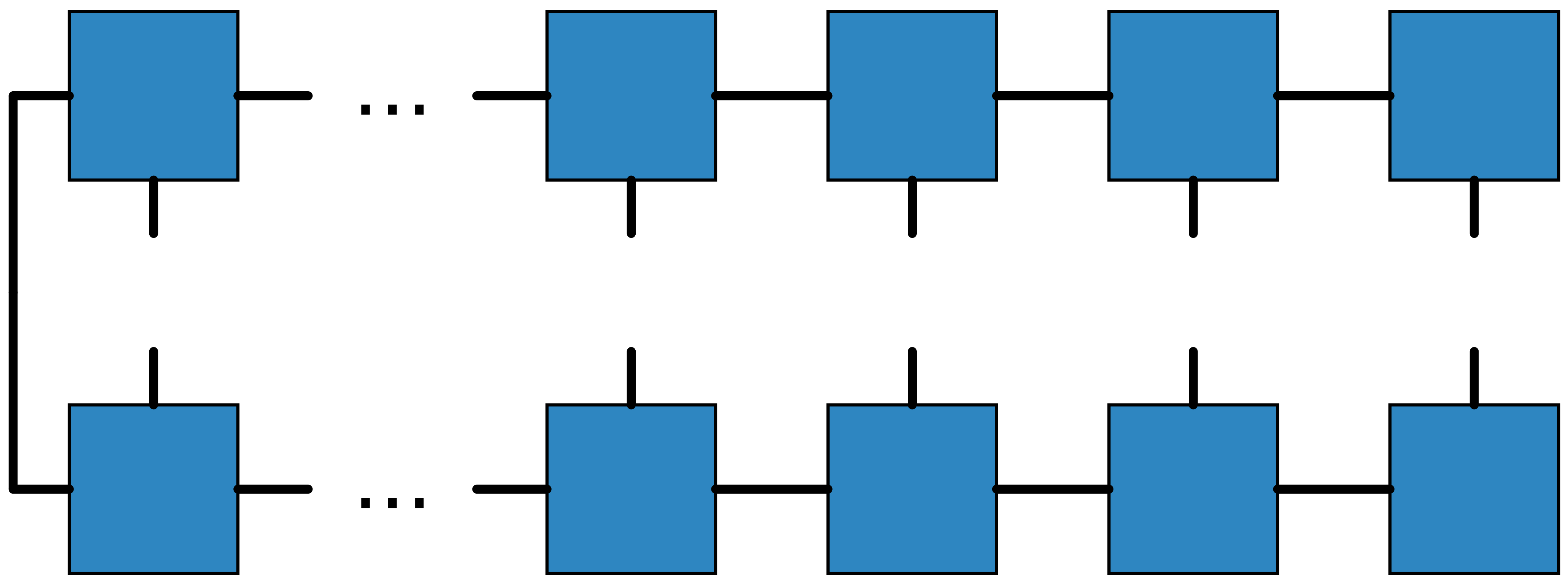
No auto-correlation compared to
slow mixing Gibbs sampling of Boltzmann Machines

Advantage-III: Direct Generation

$$p(\boldsymbol{x}) = \prod_i \frac{p(\boldsymbol{x}_{<i+1})}{p(\boldsymbol{x}_{<i})}$$

Ferris & Vidal 2012

$$p(\boldsymbol{x}_{<i+2}) =$$



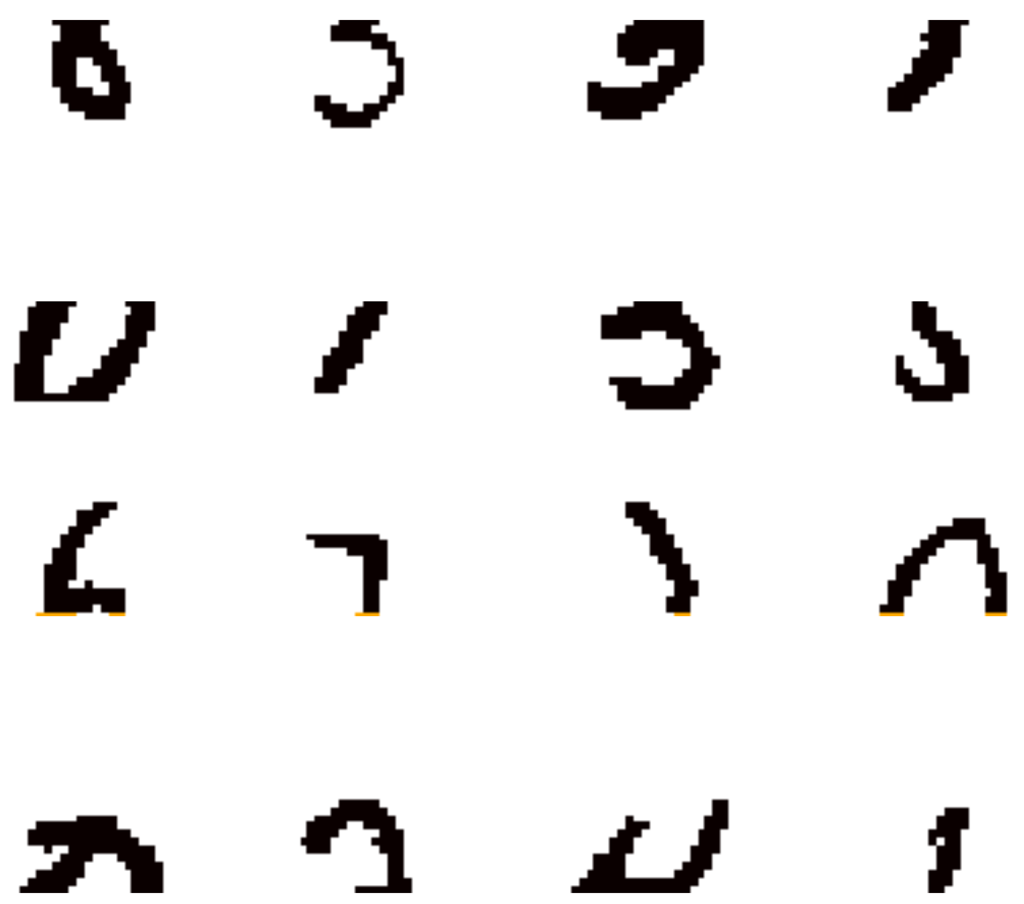
"Zipper Sampling"

No auto-correlation compared to
slow mixing Gibbs sampling of Boltzmann Machines

These advantages hold true for Tree tensor networks and MERA

Image Restoration

Han, Wang, Fan, LW, Zhang, 1709.01662



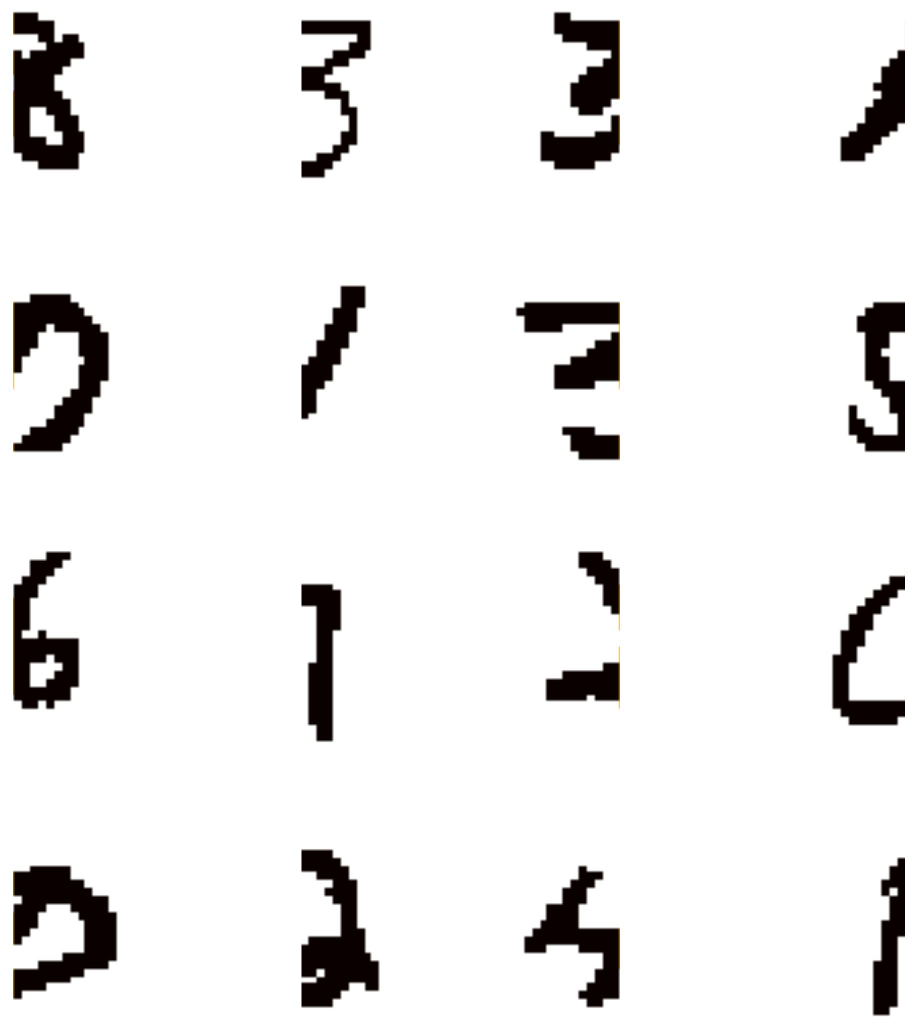
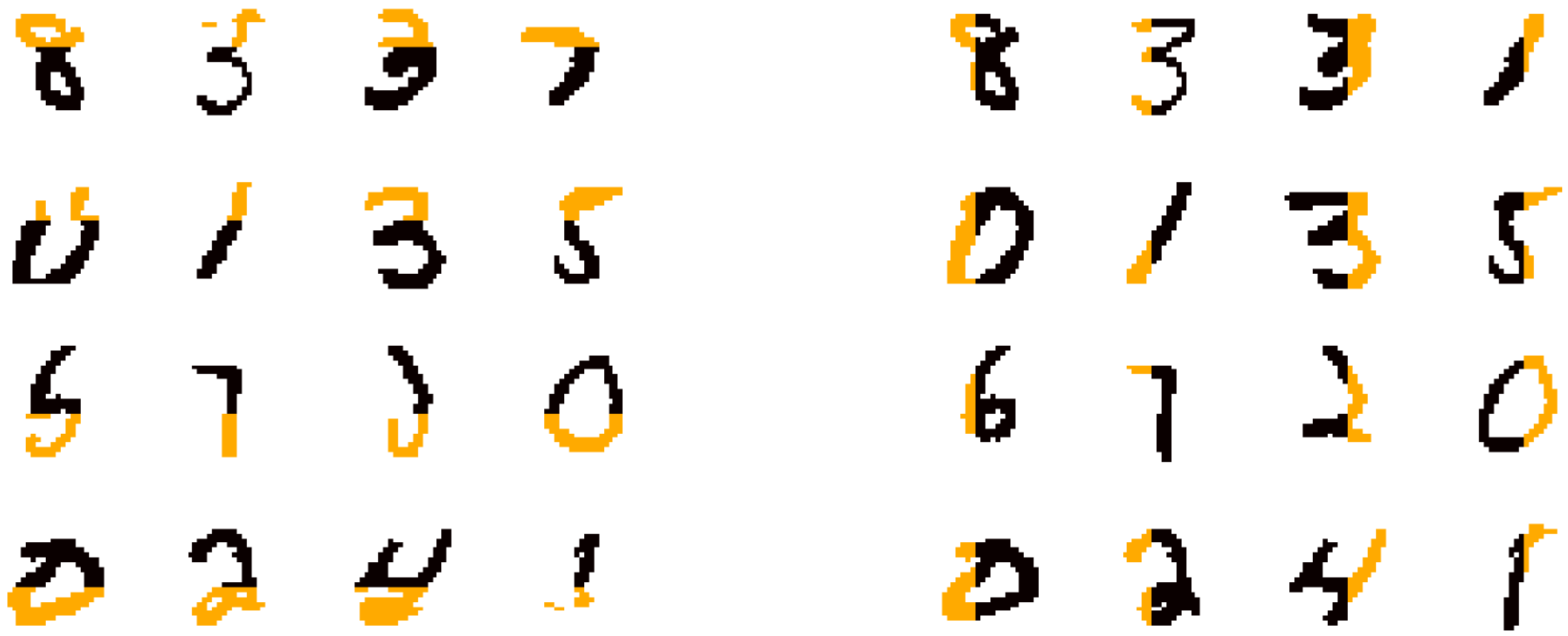


Image Restoration

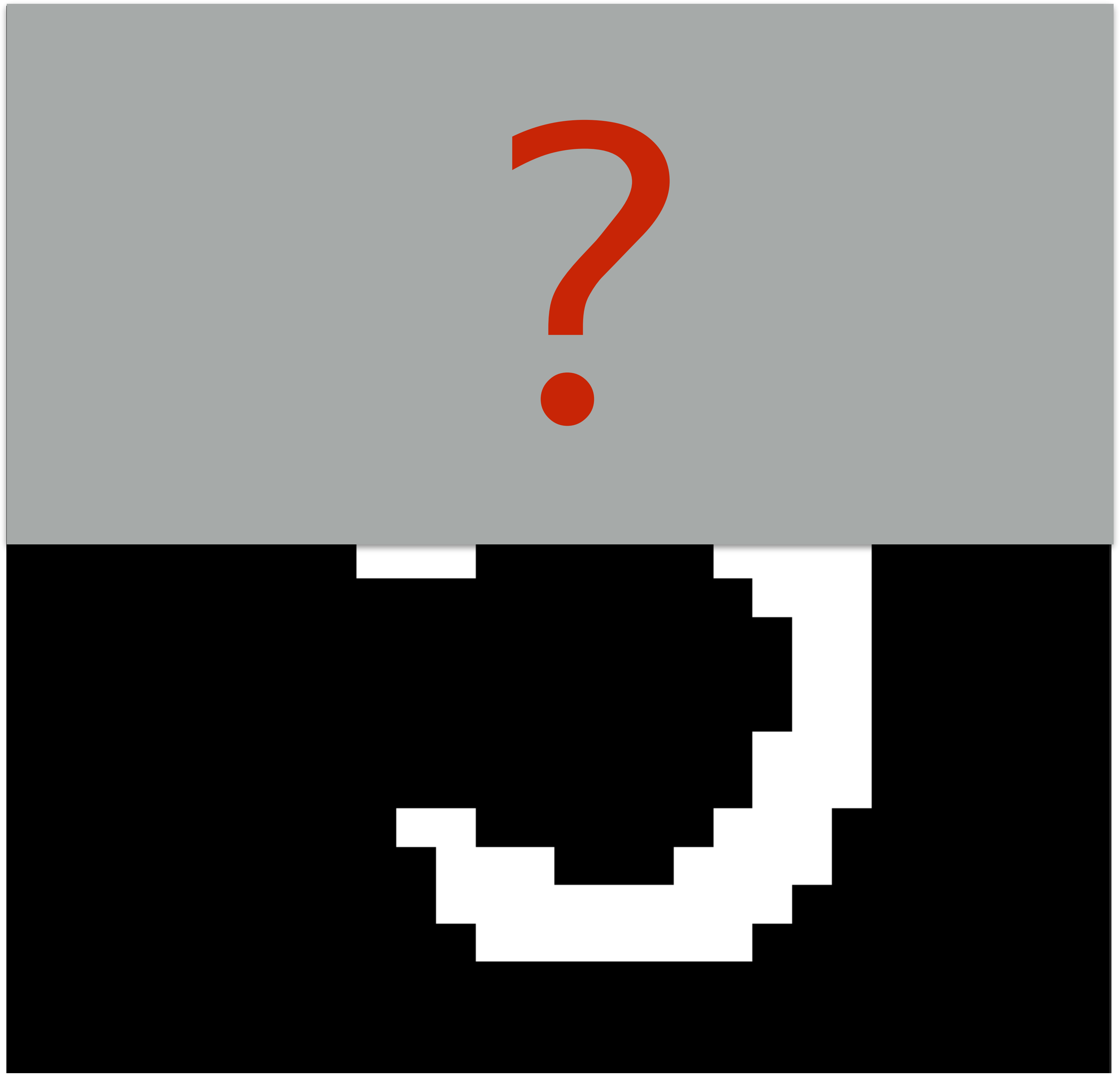
Han, Wang, Fan, LW, Zhang, 1709.01662



Arbitrary order compared to autoregressive models

PixelCNN
PixelRNN

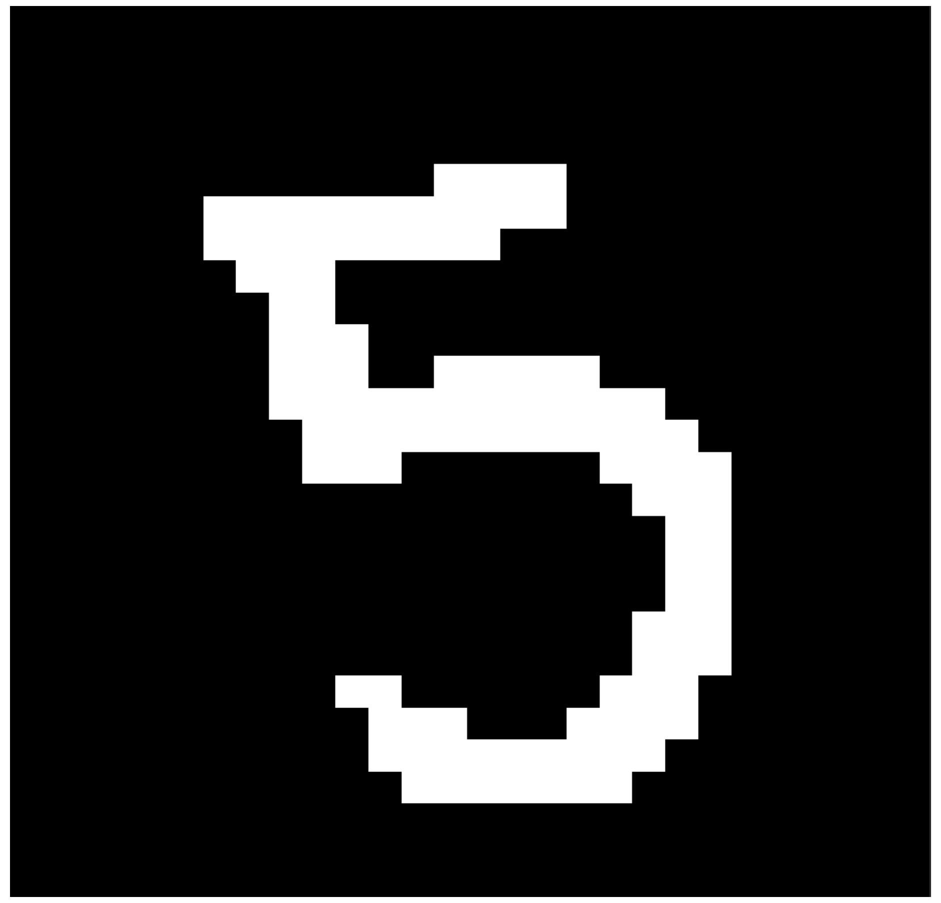
Quantum Perspective on Deep Learning

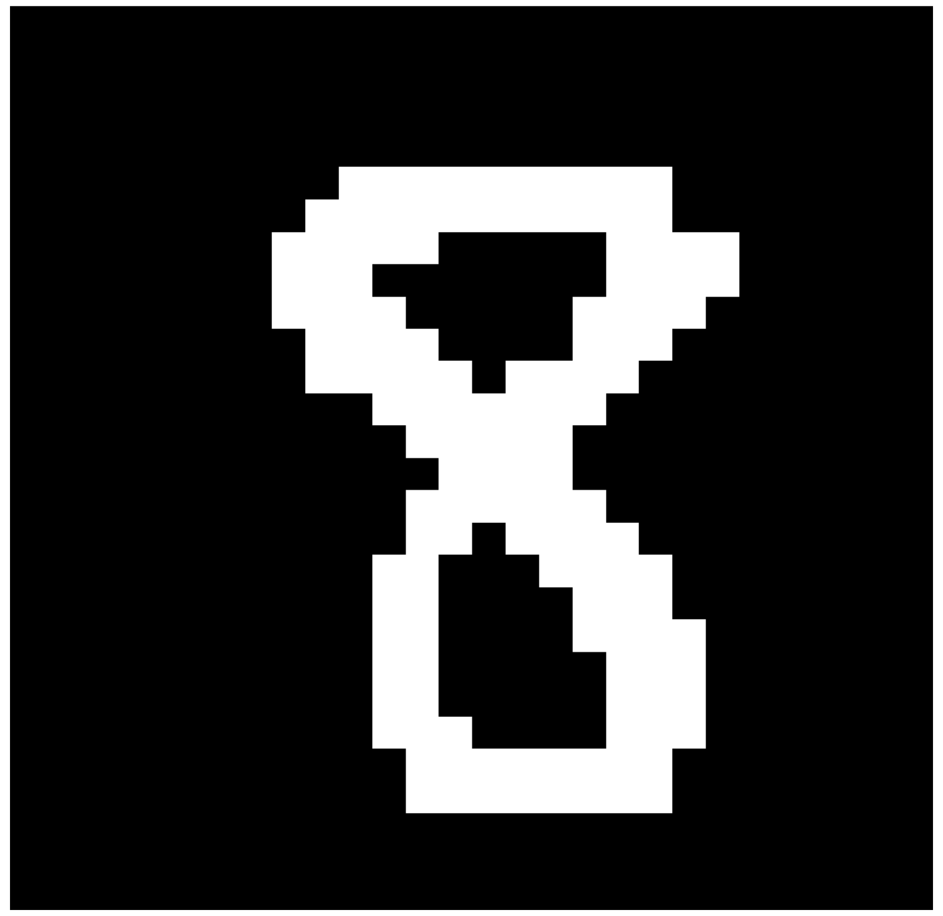


Quantum Perspective on Deep Learning

Q: How to quantify our prior knowledge on the data distribution?

A: Information pattern of the target probability functions





Quantum Perspective on Deep Learning

Q: How to quantify our prior knowledge on the data distribution?

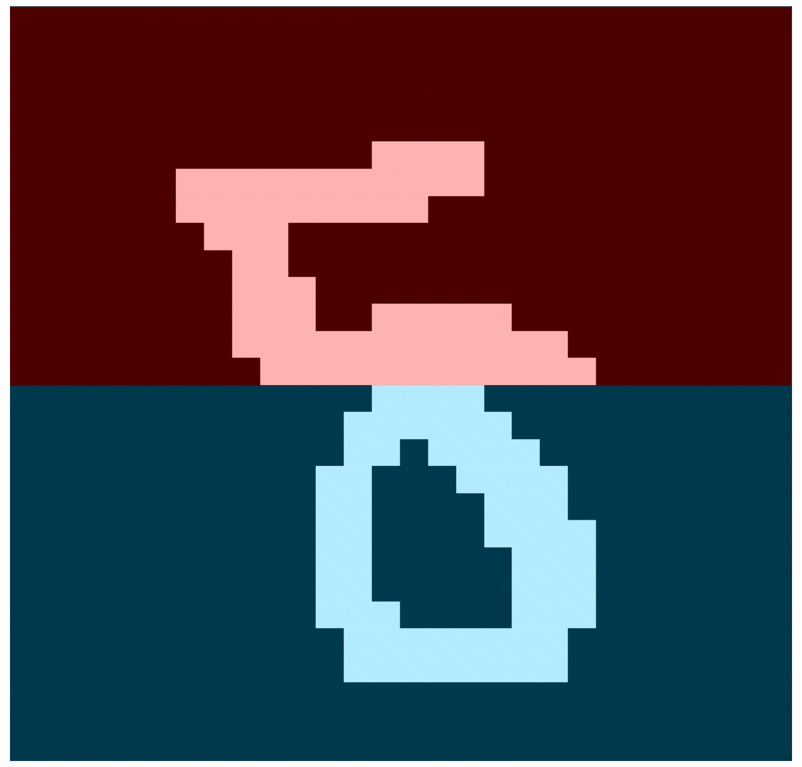
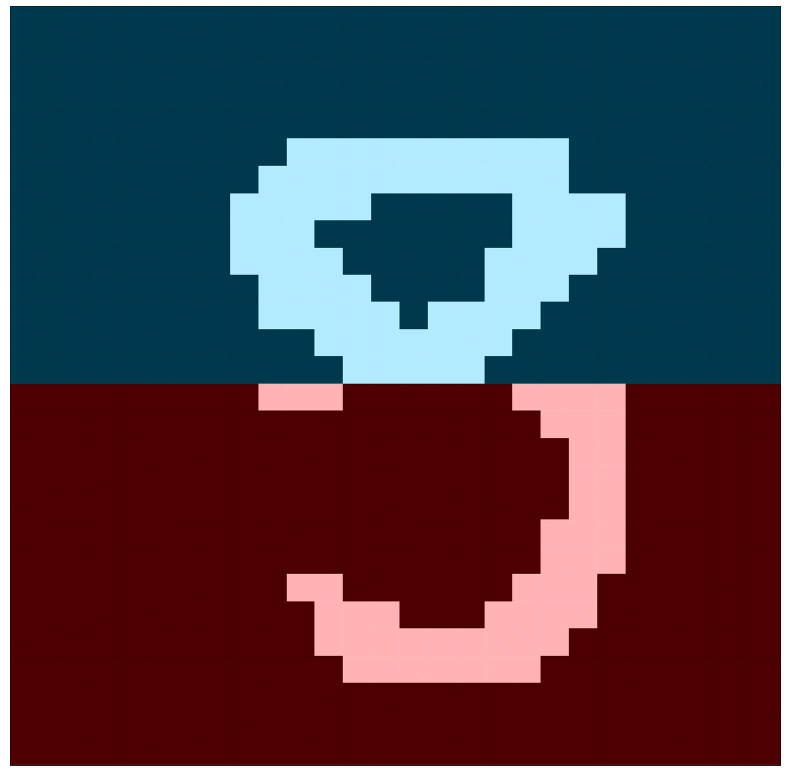
A: Information pattern of the target probability functions

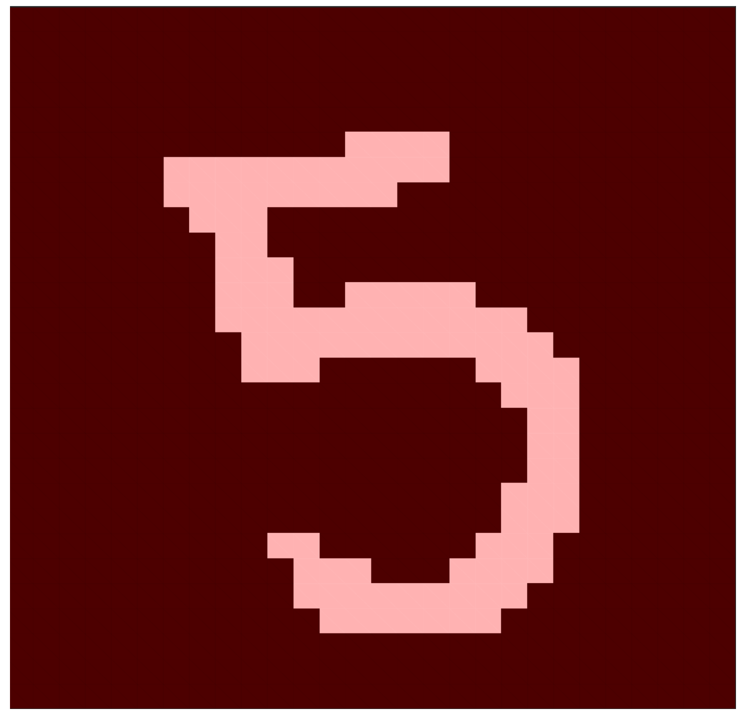
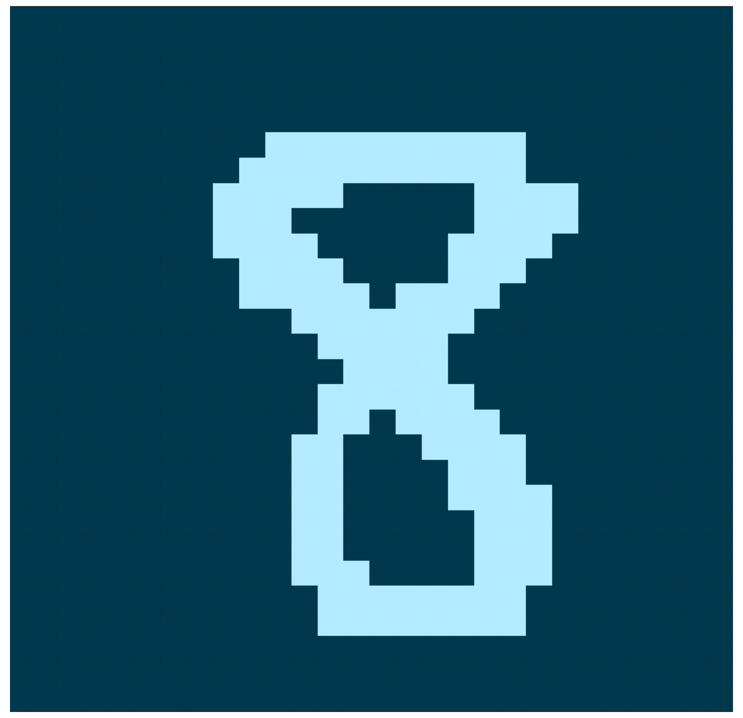
Quantum Perspective on Deep Learning

Q: How to quantify our prior knowledge on the data distribution?

A: Information pattern of the target probability functions

Quantum Perspective on Deep Learning

$p\Big($  $\Big) \times p\Big($  $\Big)$

$p\Big($  $\Big) \times p\Big($  $\Big)$

Quantum Perspective on Deep Learning

Classical mutual information

$$I = -\left\langle \ln \left\langle \frac{p(\mathbf{x},\mathbf{y}')p(\mathbf{x}',\mathbf{y})}{p(\mathbf{x}',\mathbf{y}')p(\mathbf{x},\mathbf{y})} \right\rangle_{\mathbf{x}',\mathbf{y}'} \right\rangle_{\mathbf{x},\mathbf{y}}$$

Quantum Renyi entanglement entropy

$$S = -\ln \left\langle \left\langle \frac{\Psi(\mathbf{x},\mathbf{y}')\Psi(\mathbf{x}',\mathbf{y})}{\Psi(\mathbf{x}',\mathbf{y}')\Psi(\mathbf{x},\mathbf{y})} \right\rangle_{\mathbf{x}',\mathbf{y}'} \right\rangle_{\mathbf{x},\mathbf{y}}$$

Striking similarity implies common inductive bias

+Quantitative & interpretable approaches
+Principled structure design & learning

Cheng, Chen, LW, 1712.04144

Deep Learning and Quantum Entanglement: Fundamental Connections with Implications to Network Design

Yoav Levine YOAVLEVINE@CS.HUJI.AC.IL
David Yakira DAVIDYAKIRA@CS.HUJI.AC.IL
Nadav Cohen COHENNADAV@CS.HUJI.AC.IL
Amnon Shashua SHASHUA@CS.HUJI.AC.IL
The Hebrew University of Jerusalem

Abstract

Deep convolutional networks have witnessed unprecedented success in various machine learning applications. Formal understanding on what makes these networks so successful is gradually unfolding, but for the most part there are still significant mysteries to unravel. The inductive bias, which reflects prior knowledge embedded in the network architecture, is one of them. In this work, we establish a fundamental connection between the fields of quantum physics and deep learning. We use this connection for asserting novel theoretical observations regarding the role that the number of channels in each layer of the convolutional network fulfills in the overall inductive bias. Specifically, we show an equivalence between the function realized by a deep convolutional arithmetic circuit (ConvAC) and a quantum many-body wave function, which relies on their common underlying tensorial structure. This facilitates the use of quantum entanglement measures as well-defined quantifiers of a deep network's expressive ability to model intricate correlation structures of its inputs. Most importantly, the construction of a deep convolutional arithmetic circuit in terms of a Tensor Network is made available. This description enables us to carry a graph-theoretic analysis of a convolutional network, tying its expressiveness to a min-cut in the graph which characterizes it. Thus, we demonstrate a direct control over the inductive bias of the designed deep convolutional network via its channel numbers, which we show to be related to the min-cut in the underlying graph. This result is relevant to any practitioner designing a convolutional network for a specific task. We theoretically analyze convolutional arithmetic circuits, and empirically validate our findings on more common convolutional networks which involve ReLU activations and max pooling. Beyond the results described above, the description of a deep convolutional network in well-defined graph-theoretic tools and the formal structural connection to quantum entanglement, are two interdisciplinary bridges that are brought forth by this work.

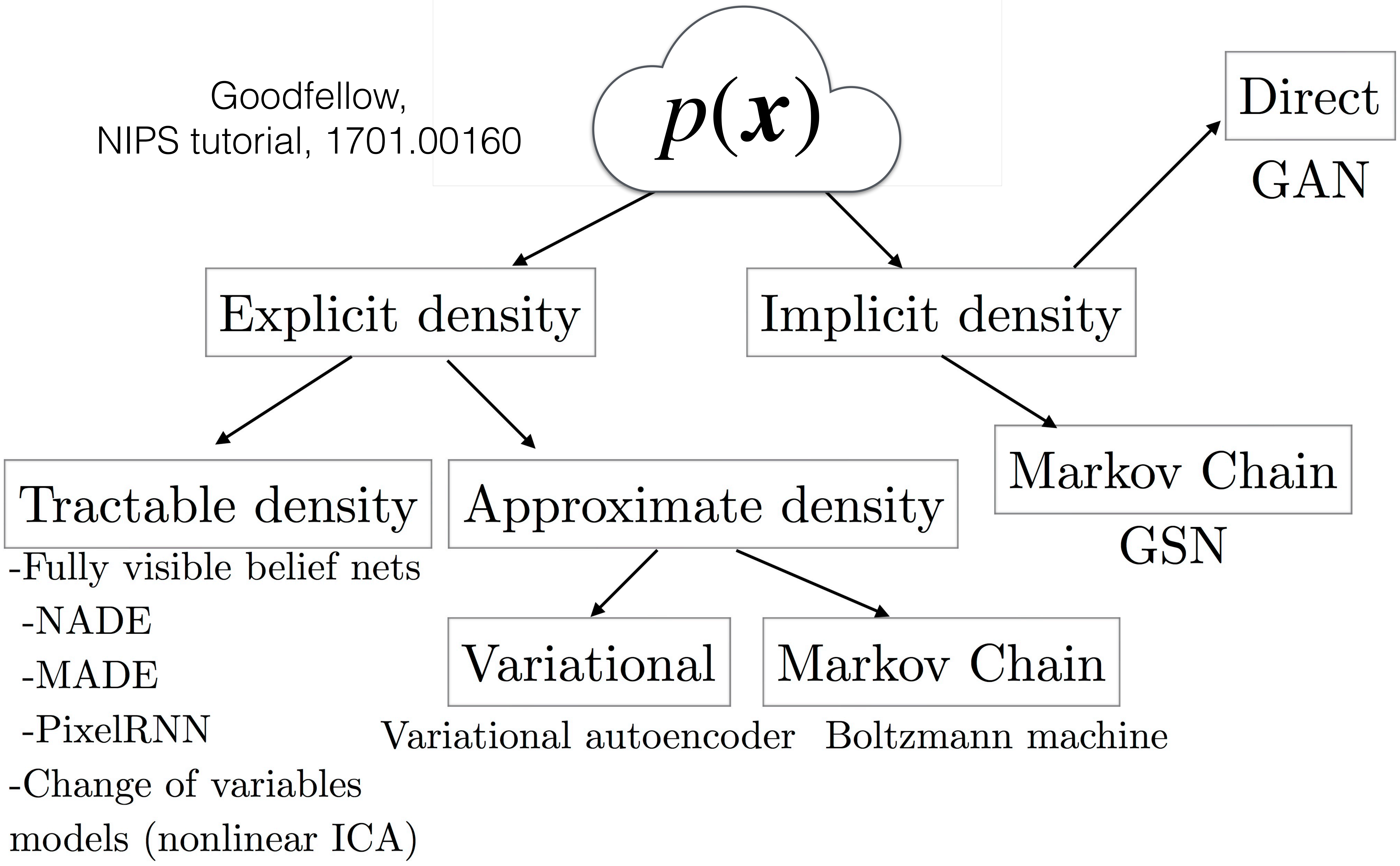
Deep Learning and Quantum Entanglement: Fundamental Connections with Implications to Network Design

Yoav Levine YOAVLEVINE@CS.HUJI.AC.IL
David Yakira DAVIDYAKIRA@CS.HUJI.AC.IL
Nadav Cohen COHENNADAV@CS.HUJI.AC.IL
Amnon Shashua SHASHUA@CS.HUJI.AC.IL

our findings on more common convolutional networks which involve ReLU activations and max pooling. Beyond the results described above, the description of a deep convolutional network in well-defined graph-theoretic tools and the formal structural connection to quantum entanglement, are two interdisciplinary bridges that are brought forth by this work.

Physics genes of generative modes

Goodfellow,
NIPS tutorial, 1701.00160



Physics genes of generative modes

Goodfellow,
NIPS tutorial, 1701.00160

$p(\boldsymbol{x})$

Explicit density

Implicit density

Direct

GAN

Tractable density

- -Fully visible belief nets
  - -NADE
  - -MADE
  - -PixelRNN
- -Change of variables models (nonlinear ICA)

Approximate density

Markov Chain

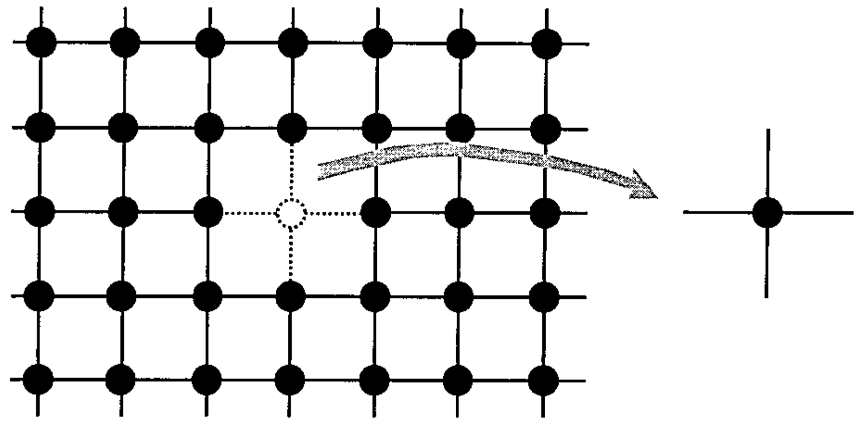
GSN

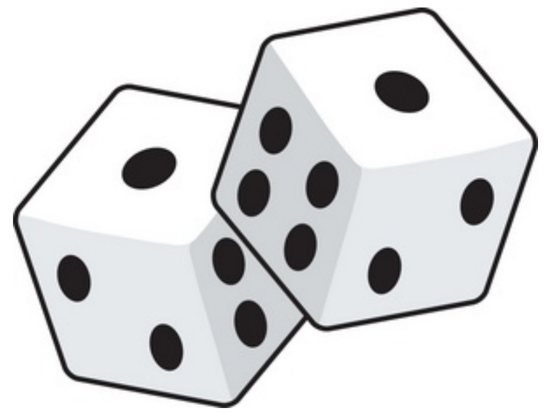
Variational

Variational autoencoder

Markov Chain

Boltzmann machine





Physics genes of generative modes

Goodfellow,
NIPS tutorial, 1701.00160

$p(\boldsymbol{x})$

Explicit density

Implicit density

Direct

GAN

Tractable density

-Fully visible belief nets

-NADE

-MADE

-PixelRNN

-Change of variables models (nonlinear ICA)

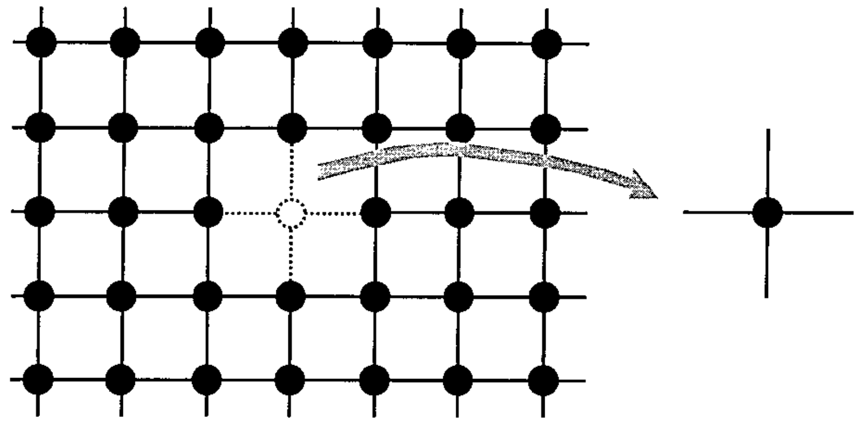
Approximate density

Markov Chain

GSN

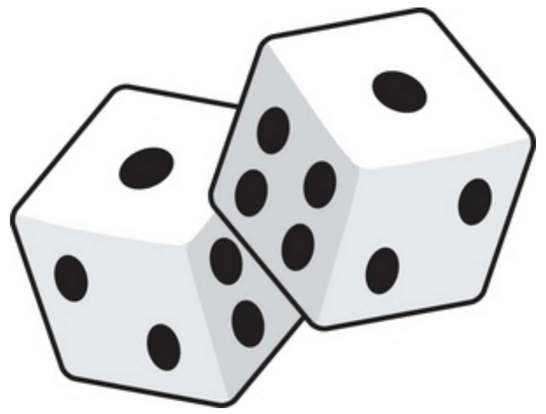
Variational

Variational autoencoder



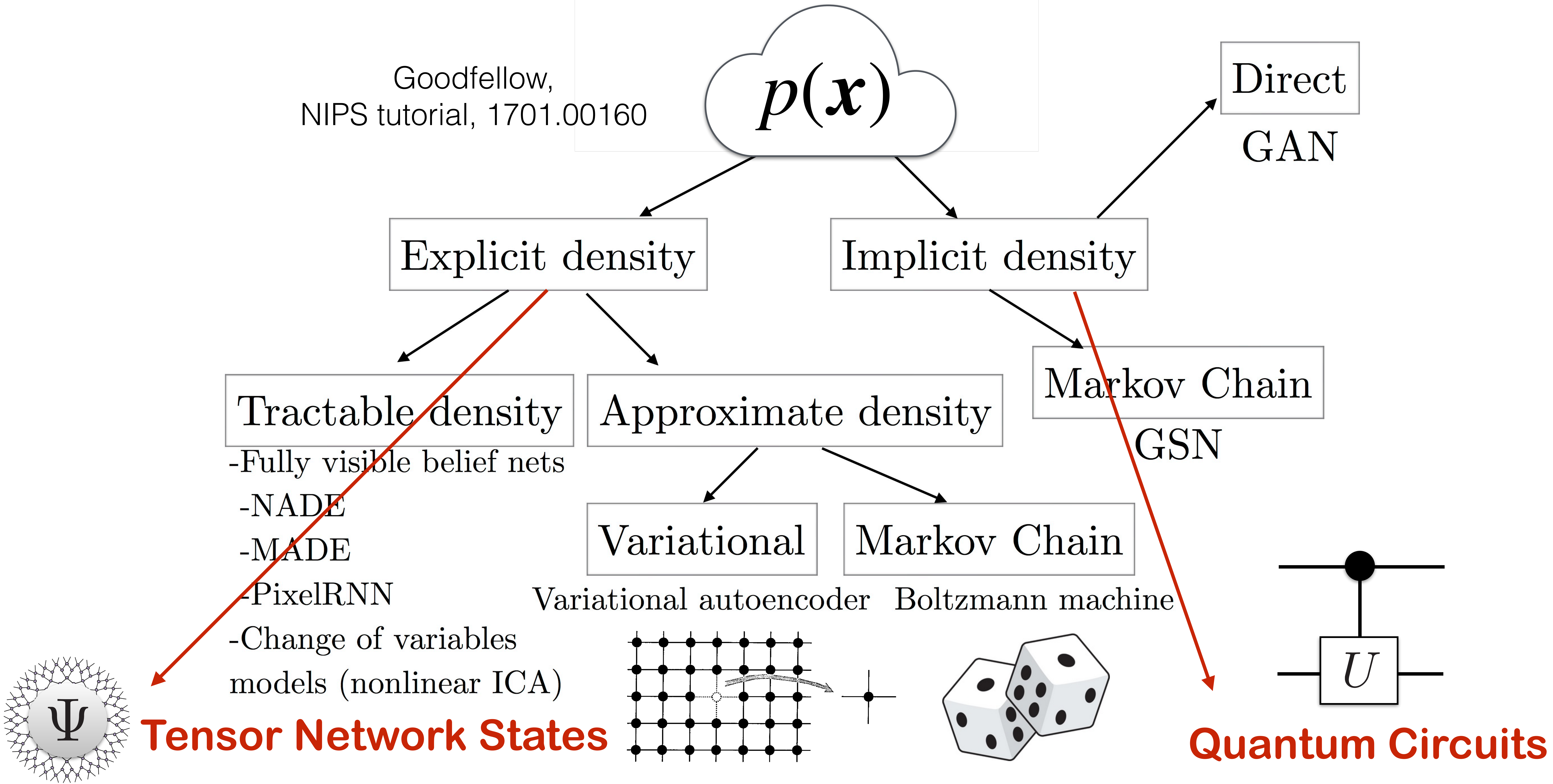
Markov Chain

Boltzmann machine



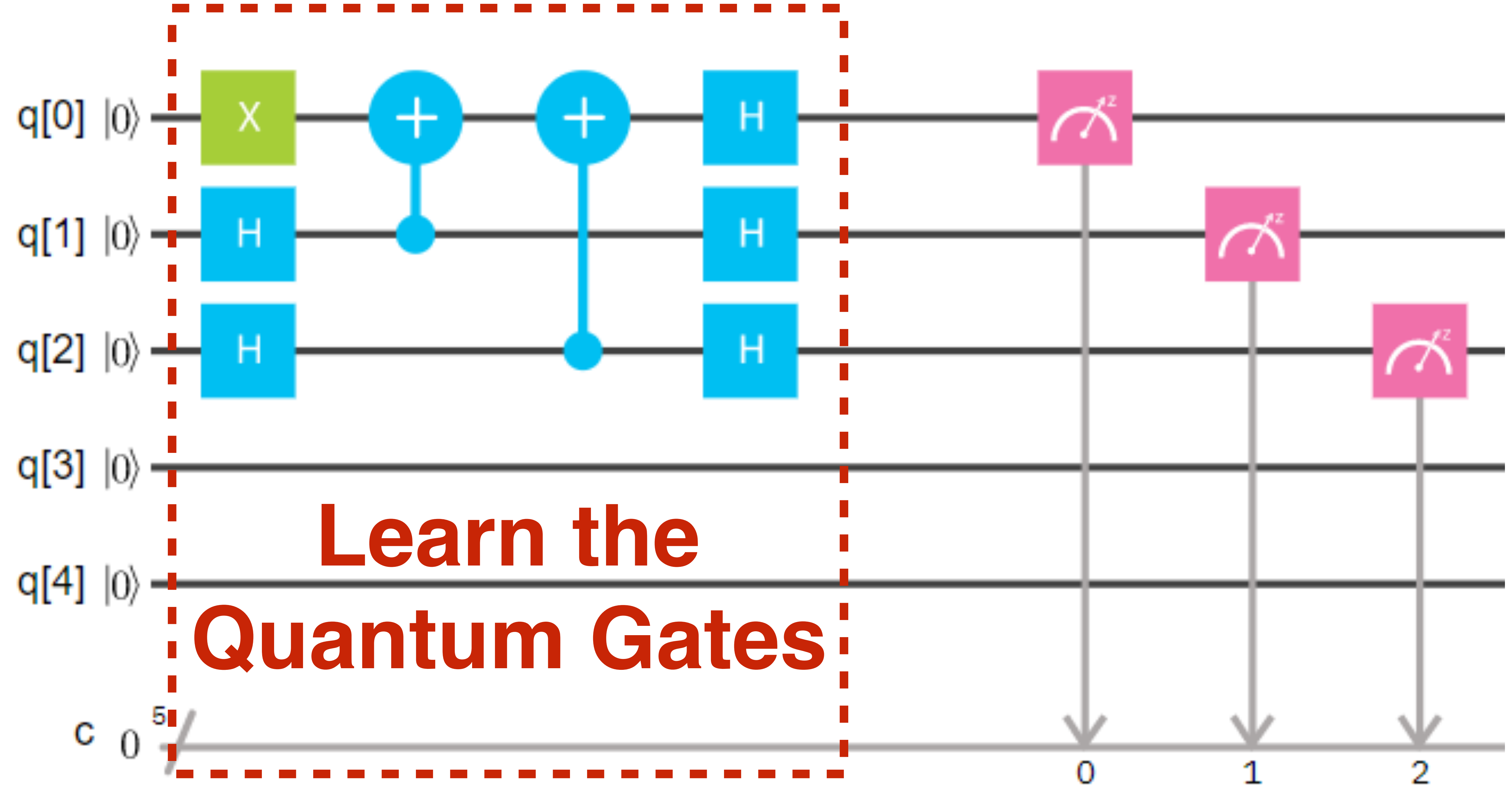
Tensor Network States

Physics genes of generative modes



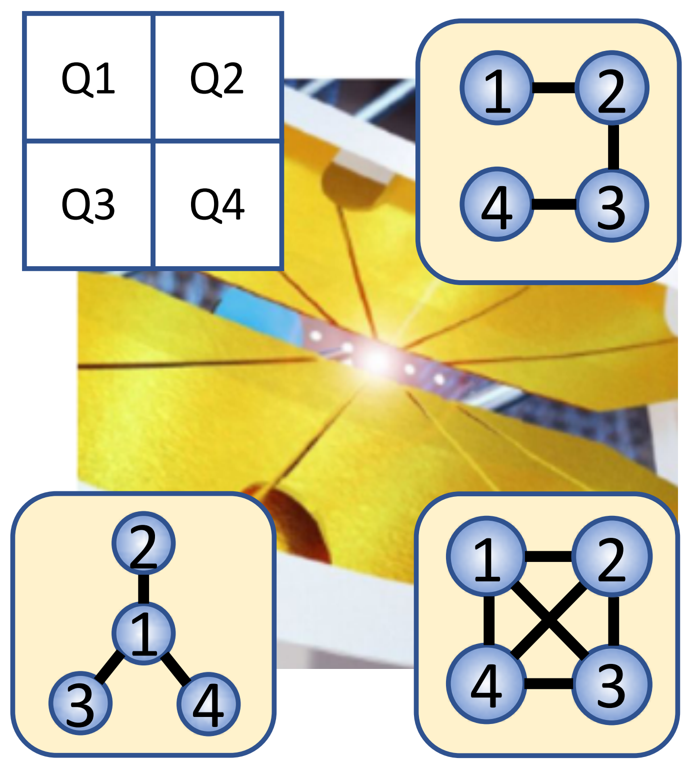
Born Machine in the cloud

Initial state

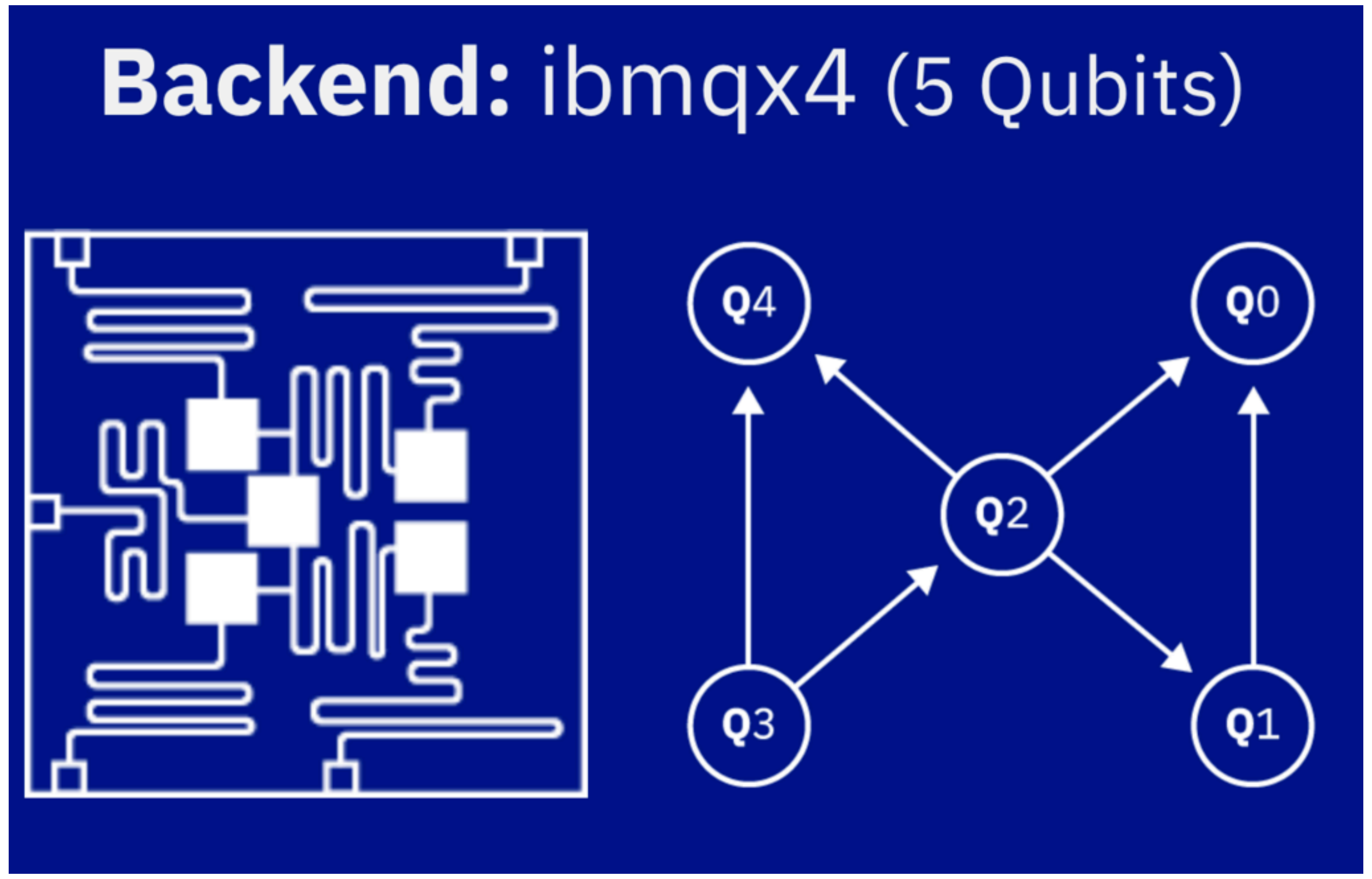


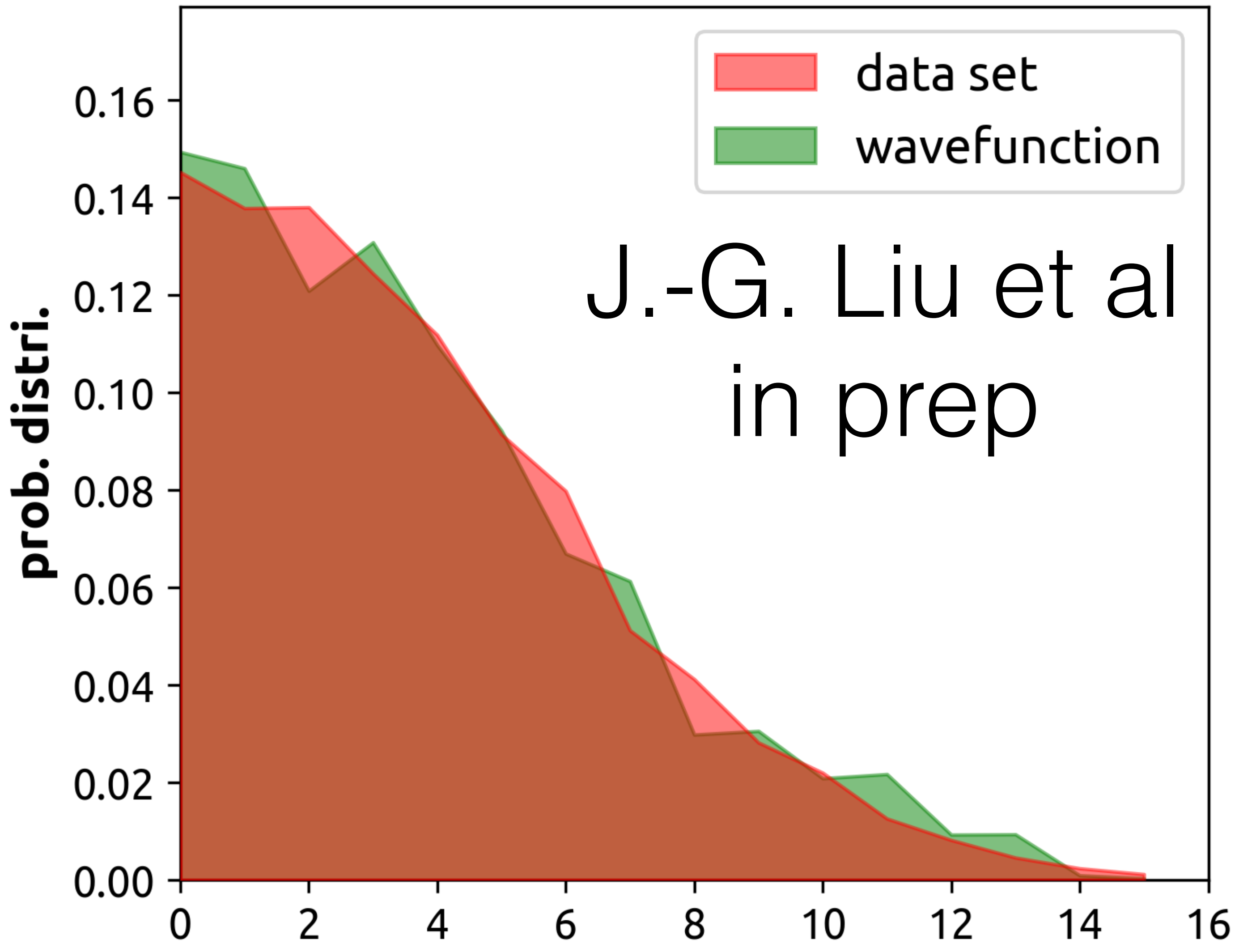
Collapse to a bit-string by measurement

Classical hard due to quantum sampling complexity

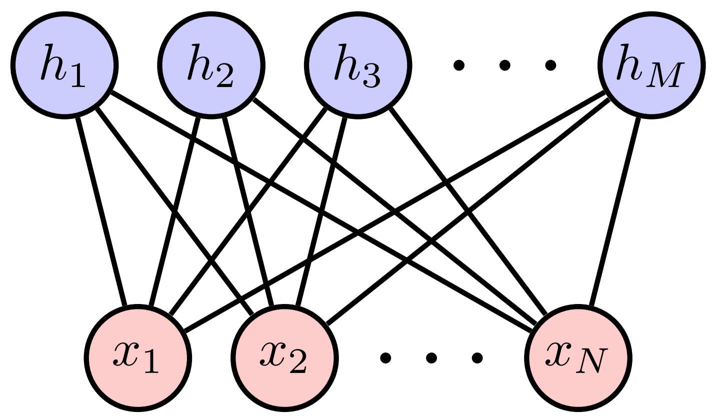


Benedetti, Garcia-Pintos, Nam, Perdomo-Ortiz, 1801.07686





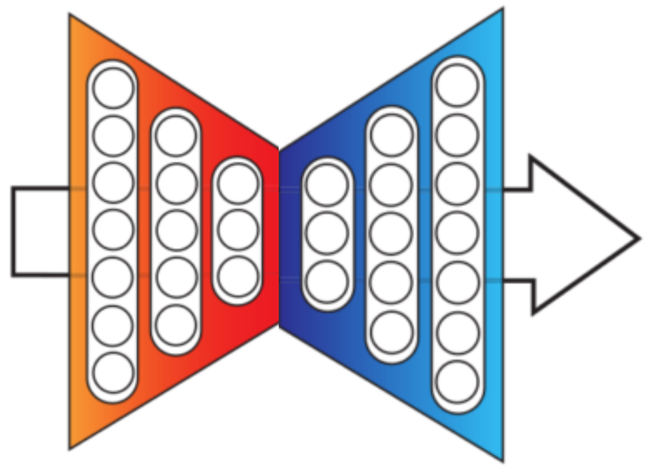
Timeline of Generative Models



Boltzmann Machines

1980s

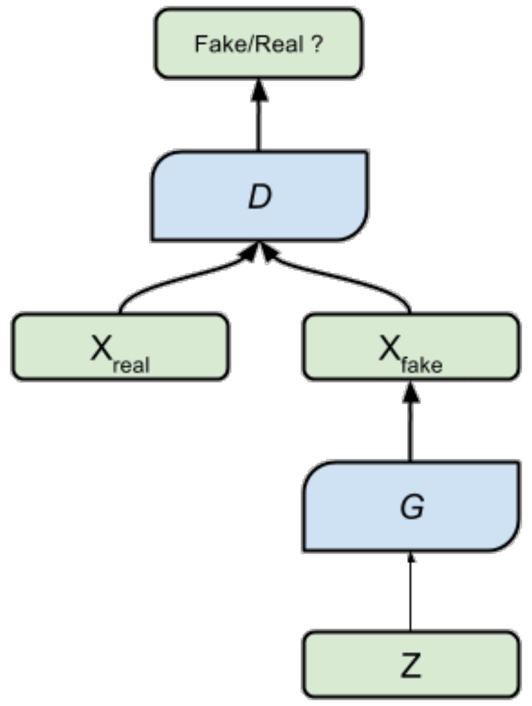
Wavefunction
Decoder
Tomography
Monte Carlo
……



Variational Autoendoer

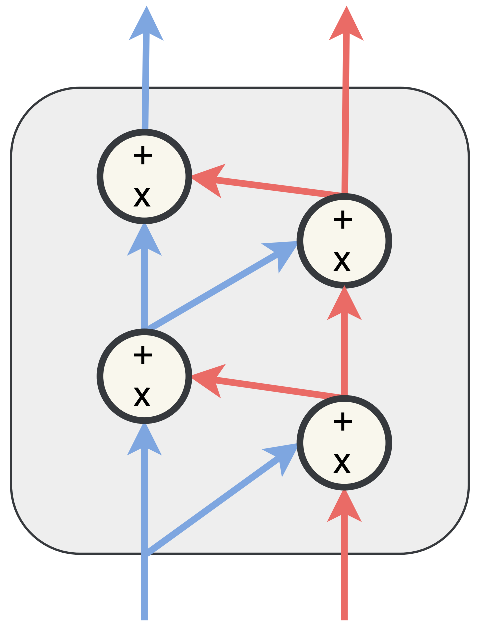
2013

Chemical
Discovery
1610.02415



Adversarial Network

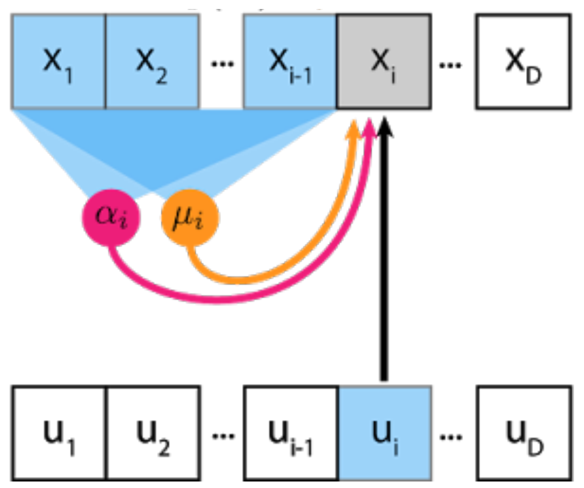
2014



Normalizing Flows

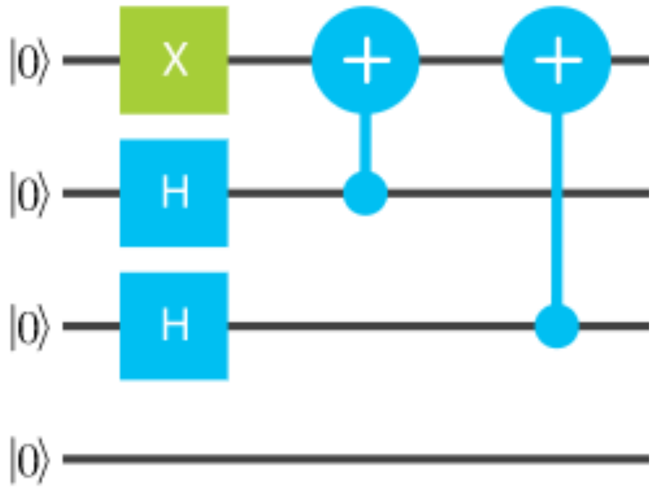
2015

Renormalization
Group
1802.02840



Autoregressive Flows

2016



Born Machines

2017

Tensor networks
Quantum circuits
Quantum perspective
Quantum Boltzmann Machines
……

Challenges ahead

- Structural learning of neural/tensor nets and quantum circuits
- Likelihood free & gradient free optimization of quantum circuits, how to scale up?
- Killer problem distributions where quantum really helps

Challenges ahead

- Structural learning of neural/tensor nets and quantum circuits
- Likelihood free & gradient free optimization of quantum circuits, how to scale up?
- Killer problem distributions where quantum really helps

Thank You!

Collaborators

Li Huang, Pan Zhang
Zhao-Yu Han, Jun Wang
Song Cheng, Jing Chen
Jin-Guo Liu, Shuo-Hui Li
Yi-feng Yang, Tao Xiang

Discussions

Zi Cai, Yehua Liu
Yang Qi, Zi-Yang Meng
Junwei Liu, Yi-Zhuang You
Xun Gao, A. Perdomo-Ortiz
G. Carleo, E.M. Stoudenmire

Neural/Tensor nets, Q-circuits visualization
https://github.com/GiggleLiu/viznet

viznet

http://wangleiphy.github.io/lectures/DL.pdf

Google Colab
free GPU access

Lecture Note on Deep Learning and Quantum Many-Body Computation

Jin-Guo Liu, Shuo-Hui Li, and Lei Wang*

Institute of Physics, Chinese Academy of Sciences
Beijing 100190, China

February 14, 2018

Abstract

This note introduces deep learning from a computational quantum physicist's perspective. The focus is on deep learning's impacts to quantum many-body computation, and vice versa. The latest version of the note is at http://wangleiphy.github.io/. Please send comments, suggestions and corrections to the email address in below.

* wanglei@iphy.ac.cn

CONTENTS

http://wangleiphy.github.io/lectures/DL.pdf

Google Colab
free GPU access

Lecture Note on Deep Learning and Quantum Many-Body Computation

Jin-Guo Liu, Shuo-Hui Li, and Lei Wang*

Institute of Physics, Chinese Academy of Sciences
Beijing 100190, China

February 14, 2018

Abstract

This note introduces deep learning from a computational quantum physicist's perspective. The focus is on deep learning's impacts to quantum many-body computation, and vice versa. The latest version of the note is at http://wangleiphy.github.io/. Please send comments, suggestions and corrections to the email address in below.

* wanglei@iphy.ac.cn

CONTENTS